DOCUMENT RESUME

ED 299 144 SE 049 669

AUTHOR	Knapp, Michael S.; And Others
TITLE	Pilot Assessment of the National Science Foundation's Investments in Informal Science Education. An Approach to Assessing Initiatives in Science Education: Volume 2.
INSTITUTION	SRI International, Menlo Park, Calif.
SPONS AGENCY	National Science Foundation, Washington, D.C. Directorate for Science and Engineering Education.
PUB DATE	Apr 88
CONTRACT	NSF-SPA-8651540
NOTE	171p.; For volume 1 see SE 049 668, summary report see SE 049 670.
AVAILABLE FROM	SRI International, 333 Ravenswood Ave., Room B-S142, Menlo Park, CA 94025 ($16.00).
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	Case Studies; Cost Effectiveness; Decision Making; Educational Assessment; Educational Finance; *Elementary School Science; Elementary Secondary Education; Evaluation Criteria; *Evaluation Methods; Evaluation Needs; Evaluation Utilization; *Financial Policy; Financial Support; *Foundation Programs; Grants; Needs Assessment; *Nonformal Education; School Support; Science Education; *Secondary School Science
IDENTIFIERS	*National Science Foundation

ABSTRACT

This project is concerned with the development of procedures and plans for the National Science Foundation (NSF) to use in assessing its initiatives on an ongoing basis. This work is meant to complement another review of NSF's investment opportunities by estabishing a sound basis for designing and carrying out assessments aimed at any of the Foundation's investments in K-12 science education, especially the new initiatives it mounts. To make ideas about assessment concrete, and to illustrate a range of activities that are not now used extensively by NSF, a pilot test of alternative assessment procedures within one area of support--informal science education--was undertaken. In this volume the findings of seven pilot-test activities are reported. These pilot assessments were intended to illustrate how short-term assessments could help NSF deal with difficult assessment questions. The results from each pilot assessment contains numerous insights into the effects of NSF's support for informal science education. (CW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED299144

AN APPROACH TO ASSESSING INITIATIVES IN SCIENCE EDUCATION

Volume 2: Pilot Assessment of the National Science Foundation's Investments in Informal Science Education

April 1988

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Michael S. Knapp
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Prepared for:

THE NATIONAL SCIENCE FOUNDATION

NSF Contract No. SPA-8651540
SRI Project No. 1809

Prepared by:

Michael S. Knapp
Mark St. John
Andrew A. Zucker
Patrick M. Shields
Marian S. Stearns
Teresa Middleton
Debra M. Shaver

With the assistance of:

Dorothy Stewart

SRI International

SE 049 677

This report presents pilot test results from the second phase of SRI's "Assessment of Initiatives Available to the National Science Foundation (NSF) in Science Education." Complementing an earlier phase of work, in which SRI discussed opportunities for the Foundation to invest strategically in K-12 science education, the second phase concentrated on ways for NSF to assess its support for science education on an ongoing basis. Both phases are part of the Foundation's response to a congressional mandate that it seek outside assistance in developing its plans and approach to managing its investments in science education.

This volume presents full write-ups of the findings from seven pilot assessments; these pilots were undertaken to demonstrate the usefulness of short-term procedures in one area of investment. In addition, an appendix presents reconstructed dialogue from one of the pilot activities, a mini-conference on the assessment of informal science learning. Readers wishing a more general discussion of assessment issues and approach are referred to *Volume 1: Design and Organization of Assessment in the National Science Foundation*.

The results of Phase I are reported in the following three volumes:

- The *Summary Report* reviews all findings and conclusions regarding NSF's mission in K-12 science education, the opportunities for the Foundation to make a significant contribution to this level of education, and how NSF can approach these opportunities more strategically.

- *Volume 1: Problems and Opportunities* presents full discussions of NSF's mission, the problems in K-12 science education that are susceptible to NSF's influence, and the opportunities to address these problems.

- *Volume 2: Groundwork for Strategic Investment* contains extended discussions of (1) NSF's "core" functions in science education (promoting professional interchange, generating information and knowledge about science education, and supporting innovation), and (2) the basis for strategic investment. This volume also includes a discussion of study methods, a summary of NSF's 30-year history of funding in K-12 science education, and three commissioned papers (regarding NSF's role in mathematics education, computer science education, and efforts to serve minority students in science).

Any of the above volumes may be requested (at the cost of printing) from SRI International, Room B-S142, 333 Ravenswood Avenue, Menlo Park, CA 94025. ATT: Carolyn Estey. Telephone (415) 859-5109

The conclusions of this report are those of the authors and contractors and do not necessarily reflect the views of the National Science Foundation or any other agency of government.

CONTENTS

INTRODUCTION 1

Overview of Phase II Pilot Test 1

Limitations of the Pilot Assessments 2

I A LIMITED CASE STUDY: ASSESSING INVESTMENTS IN COLLABORATIVE EXHIBIT DEVELOPMENT
By Patrick M. Shields and Mark St. John 3

Highlights 5

The Exhibit Research Collaborative 6

Assessing a Complex Initiative 8

Procedure for Conducting Limited Case Study. 9

Findings 11

Implications of Findings for NSF 20

II DESCRIBING THE DOMAIN OF INVESTMENT THROUGH SYNTHESIS AND ANALYSIS OF SECONDARY DATA: A "MACRO VIEW" OF INFORMAL SCIENCE EDUCATION
By Andrew A. Zucker 23

Highlights 25

Method for Constructing the Macro View 26

Informal Science Education in Context 27

Television and Informal Science Education 32

American Museums as a Source of Informal Science Education . . . 37

References 41

CONTENTS (Continued)

III MARKET ASSESSMENT FOR A NEW INVESTMENT AREA: EXAMINING THE POTENTIAL OF VIDEOCASSETTE TECHNOLOGY AS A VEHICLE FOR INFORMAL SCIENCE LEARNING IN THE HOME

By Teresa Middleton and Andrew A. Zucker 45

Highlights 47

Methods for Pilot Market Assessment 48

Pilot Assessment Findings 49

The Utility of Market Assessments 55

References 57

IV A CROSS-PROGRAM PRINCIPAL INVESTIGATORS' MEETING: EXAMINING INVESTMENTS THAT ESTABLISH LINKAGES BETWEEN INFORMAL EDUCATION INSTITUTIONS AND THE SCHOOLS

By Michael S. Knapp 63

Highlights 65

Design of the Meeting 66

Range of Existing Linkages 69

Barriers and Entry Points 71

Caveats in the Formation of Linkages 76

Implications for the Foundation 77

V AN EXPERT MINI-CONFERENCE: EXPLORING THE ASSESSMENT OF LEARNING IN INFORMAL SCIENCE SETTINGS

By Mark St. John and Patrick M. Shields 83

Highlights 85

Design of the Meeting 86

Guidelines for Assessing Informal Science Education Learning . . 88

References 103

CONTENTS (Continued)

VI **A SYNTHESIS OF LITERATURE: ASSESSING THE INFORMAL SCIENCE LEARNING EXPERIENCE**
By Mark St. John and Patrick M. Shields 105

Background. 107

The Logic of NSF's Investments in Informal Science Learning . . . 108

Approaches to the Assessment of Informal Science Learning 115

Toward an Agenda for Assessing Informal Science Learning at the Individual Level 120

References 124

VII **RETROSPECTIVE ASSESSMENT DESIGNS: A FEASIBILITY STUDY FOR INVESTIGATING THE LONG-TERM IMPACT OF INFORMAL SCIENCE LEARNING EXPERIENCES**
By Patrick M. Shields, Mark St. John, and Debra M. Shaver . . . 127

Highlights 129

The Pilot Study Approach 130

Patterns of Influence 133

Lessons Learned About Retrospective Assessments 139

References 142

Appendix: Exploring the Assessment of Learning in Informal Science Education Settings--Reconstructed Dialogue from the Meeting 145

LIST OF TABLES

Table III-1. Videocassettes Available in Science Subject Categories . . . 53

Table IV-1. Participants in a Meeting to Examine Investments Linking Informal Science Education Institutions and the Schools . . . 68

Table IV-2. How Current SEE Programs Can Support Investments That Establish Linkages Between Informal Science Education Institutions and the Schools . . . 79

Table V-1. Participants in Expert Mini-Conference on the Assessment of Informal Science Learning . . . 87

Table V-2. Goals of Informal Science Learning, as Outcomes of an Acculturation Process . . . 91

Table VI-1. Approaches to the Assessment of Informal Science Learning . . . 117

LIST OF FIGURES

Figure II-1. Americans' Use of Their Time. . . . 30

Figure II-2. Television Viewing: Who Watches How Much . . . 34

Figure III-1. Content of Existing Videocassettes . . . 52

Figure VI-1. The Logic of NSF's Investments in Informal Science Learning . . . 110

LIST OF EXHIBITS

Exhibit III-1. List of Companies to Provide Information for Assessment of the Home Videocassette Market for Informal Science and Mathematics Learning . . . 58

Exhibit III-2. Suggested Questions To Be Asked of Participants in Market Assessment . . . 61

Exhibit III-3. Outline for Market Assessment Report . . . 62

Exhibit VII-1. The Interview Protocol Used in the Feasibility Study . . . 143

INTRODUCTION

In this volume we report the findings of seven pilot test activities, each of which was designed to demonstrate a way of assessing initiatives of the National Science Foundation (NSF) in informal science education. These pilot assessments are not complete studies; rather, they should be thought of as feasibility tests, intended to illustrate how short-term focused assessments could help NSF deal with difficult assessment questions. Nonetheless, the write-up of results from each pilot assessment contains numerous insights into the effects of NSF's support for informal science education.

This volume will be of interest principally to those who wish to know more about NSF's investments in the informal science education domain, or who wish to understand the methodology of the pilot test activities in some detail. Readers interested in more general questions about assessing NSF's investments, including the application of these pilot activities to other areas of science education support, are referred to Volume 1 of this report, *Designing and Organizing Assessment in the National Science Foundation*. Sections VII-IX of that volume briefly describe the pilot studies by reviewing their purposes, methodology, illustrative findings, and considerations for further application of the procedures.

Overview of Phase II Pilot Test

Phase II of the SRI project "An Assessment of Initiatives Available to the National Science Foundation in K-12 Science Education" is concerned with the development of procedures and plans for NSF to use in assessing its initiatives on an ongoing basis. This work is meant to complement SRI's review of NSF's investment opportunities carried out in Phase I, by establishing a sound basis for designing and carrying out assessments aimed at any of the Foundation's investments in K-12 science education, especially the new initiatives it mounts.

To make our ideas about assessment concrete, and to illustrate a range of activities that are not now used extensively (or at all) by NSF, we undertook a pilot test of alternative assessment procedures within one area of support: informal science education. This domain was chosen for various reasons, among them: the Foundation's investments in informal science education are sizable; NSF has a long track record of investment in this area; the domain represents an expanding area of investment with promising new possibilities; the diversity of investments allowed us to try out a wide array of procedures; and investments in this domain raise some of the most difficult, interesting assessment issues of any area in NSF's portfolio.

We recognized that by confining our pilot test to a domain that differs significantly from investments aimed at formal schooling it may not be so easy to see the applications of these ideas to other areas. In Volume 1 we have tried to suggest considerations affecting the application of these procedures to a variety of areas in science education.

Limitations of the Pilot Assessments

In reading and interpreting the results of these pilot studies, the reader needs to keep in mind the following caveats:

- The assessments were designed and carried out to illustrate a procedure rather than to produce definitive findings.

- All of the pilot assessments were carried out simultaneously under time and resource constraints that might not pertain if NSF had undertaken these studies as part of its normal assessment routine.

- Even if they had been conducted as complete studies, the pilot assessments would not constitute a full evaluation of any given initiative. Rather they are intended to answer a focused set of assessment questions about particular initiatives, areas of investment, or problems of assessment design. The answers to these questions are one of many sources of information necessary to make informed judgments about the Foundation's support for science education (see discussion of complementary information sources in Sections II, V, and VI in Volume 1).

- Some of the pilot assessments are more complete than others. The limited case study of the Exhibit Research Collaborative (see Section I), for example, rests on a sound qualitative data base; its results would probably be replicated by more intensive investigation of this investment. Others, such as the demonstration of retrospective techniques as a way of examining long-term influences of informal science learning (see Section VII), assembled a more meager data base; the results of this pilot should be taken as illustrative and suggestive (although they converge with findings from related research, as we point out in Section VII).

Because of these limitations, NSF decisionmakers and others should not rely on the pilot assessment findings as the sole or even primary source of information about the initiatives in question. Rather, in reading through these pilot assessments, report audiences should use the studies to improve their understanding of informal science education as an area for the Foundation to support. They can also use the findings as a way to highlight and "test" some of the assumptions underlying the Foundation's initiatives in this area. Hopefully, these pilot assessments will whet the appetite of readers for similar (although more complete) studies that span the full range of initiatives in science education supported by NSF.

I A LIMITED CASE STUDY: ASSESSING INVESTMENTS IN COLLABORATIVE EXHIBIT DEVELOPMENT

By

Patrick M. Shields
Mark St. John

I A LIMITED CASE STUDY: ASSESSING INVESTMENTS IN COLLABORATIVE EXHIBIT DEVELOPMENT

The purpose of this pilot activity was to illustrate how a limited case study of an ongoing initiative could be carried out and how it might be useful to NSF as well as to the projects studied. As the focus of the pilot activity, we selected the Exhibit Research Collaborative (ERC), a joint effort by eight primarily midsized science centers to build and share high-quality science exhibits. Although it is funded as a single project, the ERC resembles multiproject initiatives and thus afforded a manageable way to test an assessment procedure that could be used to study NSF initiatives that involve a range of projects.*

Highlights

The range and depth of our findings in this study are limited by our focus on modeling the feasibility of the brief case study procedure, by the relatively short time we spent in each site, and by the fact that the project is in progress, with a number of the science centers still in the planning stages. Nevertheless, we believe we can legitimately draw the following conclusions from our study of the ERC:

- *The collaborative mechanism appears to be working*--in the sense that it is functioning as envisioned, producing exhibits of apparently high quality that are being shared among participating institutions. To date, the most serious problems with the collaborative have been technical--e.g., involving the durability and adaptability of exhibits as they travel among institutions.

- *The collaborative relationship among participants has developed without seriously hampering the autonomy of each*; however, it has taken time and effort to establish the requisite level of trust among institutions.

- *The collaborative mechanism appears to achieve leverage of several kinds*: it provokes greater effort by museum staff to produce high-quality exhibits; it facilitates the sharing of resources among consortium members; it creates a repertoire of exhibits for medium-size museums that significantly augments their own collections; and in some cases it enhances consortium members' fund-raising capability. In these ways, this mechanism gives NSF a way to catalyze exhibit development nationwide in a category of museums it has heretofore not reached.

---

* In many respects, our analysis treats the ERC as if it were an NSF initiative and, strictly speaking, this is not the case. The collaborative is, in fact, a field-initiated project designed to meet the needs of the participating members. But for purposes of conducting this pilot test, the ERC was an ideal test case. As we note at the close of this section, NSF may wish to consider pursuing this or similar lines of investment more purposefully than it has heretofore.

- *The collaborative has encouraged and supported "formative" evaluation*, that is, structured self-assessment done by participating museums to improve the project as it proceeds. This has helped exhibit designers to match exhibits more successfully to the audiences that use them; however, without further assistance, member institutions are unlikely to institutionalize this process.
- *Each member institution benefits from, contributes to, and participates in the collaborative differently*, depending on its size (in relation to ERC exhibits), philosophy, institutional goals, political motivations, and other factors unique to each site.

The Exhibit Research Collaborative

The collaborative was formed in response to a common need of medium-size museums for high-quality, interactive traveling exhibits. These museums include institutions that fall between the large, long-established major science museums in the nation's largest urban centers and the smaller "start-up" science centers that have proliferated over the past decade. The midrange science centers share the characteristics of being relatively stable and financially secure, while still struggling to develop a critical mass of permanent high-quality exhibits. As they grow and evolve into institutions with increasing attendance and larger numbers of repeat visitors, these centers frequently find themselves without the financial resources or staff capacity to build a sufficient number of high-quality exhibits annually. Center staff, then, are forced to build additional exhibits at a pace that does not allow sufficient research and development time, or these institutions attempt to lease traveling exhibits (over which they have little quality control). In addition to leaving the center with less-than-desirable exhibits, these two approaches to meeting the demand for new exhibits provide few learning opportunities for museum staff. The centers are then caught in a "Catch-22."

To grow and to meet the demands of their increasing audience, exhibit design staffs must rush to put together or borrow new and exciting exhibits. But in the process, they are not building the institutional capacity necessary to meet the demands of subsequent years.

The ERC was designed to solve this problem by focusing each museum's energy and resources on the development of a single exhibit, which is then shared with the other participating institutions. Because staff at each center have ample time and resources to build one exhibit, they can ensure that it is of high quality. At the same time, the center has the guarantee of a stream of good traveling exhibits for a couple of years into the future. Moreover, the process of building an exhibit that must meet the standards of all the participating museums, as well as those of NSF, encourages an exhibit-building process that helps the center to develop its staff and ultimately its capacity to build exhibits in the future. The ERC also contributes to further staff development through a well-articulated process of formative evaluation in which staff at each center are encouraged to review and test the quality and educational value of the exhibits they build.

Over 3 years, NSF will contribute $1.14 million to the collaborative, while each science center will provide an additional $100,000. Using a staggered schedule, each center follows a similar design sequence that includes assessing visitors' knowledge and interests, building and testing prototype exhibits, and creating a finished copy of the exhibit, which will travel to the other museums. The first exhibits were finished in mid-1987 and are currently traveling to other centers. The last exhibits to be built will be completed in the summer of 1989. Each institution receives a traveling exhibit every 4 months for the following 28 months (seven exhibits in total). Traveling exhibits are scheduled to be on the floor of each center for 3 months, with 1 month allowed for shipping and setting up. The last exhibits built will end their travels in the spring of 1991, although some centers have already made arrangements to lease their exhibits to nonparticipating science centers after that date. Throughout the process, representatives of the various institutions meet regularly to review exhibit topics, assess the collaborative's progress, and hammer out technical details.

The participating centers are:

- Impression 5 Science Museum, Lansing, MI
- The Museum of History and Science, Louisville, KY
- The Museum of Science, Boston, MA
- The Oregon Museum of Science and Industry, Portland, OR
- The Pacific Science Center, Seattle, WA
- The Reuben Fleet Space Theater and Science Center, San Diego, CA
- Discovery Place of Charlotte, NC
- The Science Museum of Virginia, Richmond, VA

Given the similarity of many of the names, we will refer to each center by the name of the city in which it is located.

The ERC is not the first consortium of museums that have pooled their resources to "leverage" their exhibit-building efforts. The Science Museum Exhibit Collaborative (SMEC) is several years older than the ERC and includes eight of the largest science museums in the country. Science museums in the Northeast have pooled their resources in another consortium to purchase and share a very expensive dinosaur exhibition. More recently, a consortium has been formed by the omnimax theaters to promote the production and sharing of high-quality films.

The ERC, like SMEC, is engineered to allow institutions to work together in ways that are productive and that minimize interinstitutional friction. The basic philosophy rests on the assumptions that each member is competent to build an exhibit of high quality, and that each institution should have independence in the design and construction of that exhibit. Deliberately, there is no attempt to design or to build exhibits as a group or a committee.

The members of the collaborative met initially to agree on the topics that each museum would address and to review the basic idea of the exhibit design. At this time, members also specified the physical characteristics of the exhibits, as well as the schedule and logistics of the traveling. Overall, however, the collaborative built in a great deal of autonomy and independence for each institution.

Assessing a Complex Initiative

The ERC project resembles NSF initiatives that reflect a complex "hypothesis" about the effects of Foundation funding (the project may also be thought of as a prototype for a more formal initiative in the future). The project alters NSF's funding for museum exhibit development by pooling the efforts of individual proposers and by building the capacity of those institutions involved in the project. Through a process of professional collaboration and with the technical assistance of a professional evaluator, the initiative is designed to support a group of disparate institutions in building high-quality interactive exhibits that will provide educationally fruitful experiences for individual visitors in a wide variety of settings.

Because NSF's Informal Science Education program has placed a high priority on reaching large numbers of people in a cost-efficient manner, smaller science museums have had difficulty competing with large urban science centers for exhibit development funds. Traditionally, NSF has focused its resources on larger, well-established institutions that can provide exhibits and models for other museums to emulate. The ERC collaborative provides NSF with a mechanism that allows it to fund smaller science centers and still ensure high leverage. The project is nationwide, may run for over 5 years, and is intended to reach more than 40 million people. In addition to developing eight high-quality exhibits, NSF hopes that the process will build both personal and institutional capacity in a variety of forms.

Since the ERC project represents an experiment about the use of a collaborative mechanism to gain leverage and to build capacity, a brief assessment of the project, we believe, can provide NSF with some general lessons about this initiative as well as give feedback to the institutions involved. For NSF, data on the broad national effect of the project (e.g., number of visitors) and its cost-effectiveness could prove useful in reporting inside the organization and to interested outside parties such as Congress. Perhaps more important, information on the functioning of the collaborative mechanism might help NSF planners to refine projects of this type in the future or even to assist current ERC members. For the participating institutions, information on the progress of the collaborative might assist them to make midcourse corrections or to plan future endeavors.

Assessment of the ERC must take into account several important characteristics of the project. First, it is a multisite project, with each participating science center having its unique culture, educational philosophy, political agendas, and reasons for wanting to participate in the collaborative. Second, the project has a number of goals: developing good exhibits, reaching large numbers of people, implementing a formative evaluation component, etc. Third, because the project is in midcycle, any information collected can only approximate the potential outcomes of the project. Finally, the type of data to be collected is quite imprecise: what was the progress of the formative evaluation process? How did the prototype testing affect the final design of the exhibit? Are the exhibits any good? Are there any institutional effects on the participating museums?

Given these characteristics, it is necessary to carry out an assessment using an approach that is open-ended, exploratory, and sensitive to a variety of ill-defined outcomes. One appropriate and effective way of collecting this type of data is a multisite case study that allows for on-site, in-depth analysis using multiple methods of data collection (e.g., interviews, observation, record review). Traditional methods of assessing the progress of projects (e.g., annual project evaluation reports, principal investigator meetings, ad hoc telephone conversations) would not provide the depth of information that full case studies yield. For example, if one were interested in gauging the effect of the formative evaluation process on the design of an exhibit, it would be more useful to sit down with an exhibit designer and review actual prototypes, design plans, and the finished exhibit to chart its changes over time and compare these with results of evaluation efforts than it would be to simply read a text describing the process.

Unfortunately, the cost of doing full case studies at each site would be high. Consequently, in the pilot test, we attempted a modified case study approach in which costs were reduced by reducing the number of sites visited and limiting time on-site. Our case study of the ERC can be seen as a test of the usefulness of limited case studies for learning about current initiatives, whether they consist of a single complex project like the ERC, or multiple projects engaged in similar endeavors.

Procedure for Conducting Limited Case Study

Our primary criterion for selecting case sites among the eight participating institutions was a center's position in the exhibit design schedule. Accordingly, we identified four separate stages in the development process and selected one institution that fit into each:

- Planning stage:	Charlotte
- Prototypes developed:	Seattle
- Final exhibit completed:	Louisville
- Received traveling exhibit:	Portland

In addition, we chose to visit two other participating centers that were not mid-size museums. The Boston Museum of Science is a large, well-established institution, quite different from the other science centers in the collaborative and, arguably, the least in need of participation in such a collaborative. In contrast, the Reuben Fleet Science Center in San Diego is considerably smaller than the other institutions, with greater fiscal, personnel, and physical constraints. Finally, we included a visit to the Science Museum of Virginia in Richmond because its staff evaluator had visited each of the participating institutions and could provide excellent background for our subsequent visits.

The sequence of our site visits was determined primarily by the scheduling constraints of the various museums and geographic considerations. We did, however, structure our schedule to begin in Richmond, in order to collect background information on each of the sites, and to end at Portland, in order to discuss our general impressions with the ERC project leader (who is located in Portland), who could help us interpret what we had learned from all the sites we visited.

Two staff members took part in the case study, although each site visit involved only one person on-site. Site visits lasted one full day and included semistructured interviews with relevant museum staff, a review of records concerning the exhibit development process, and, where possible, a physical review of prototypes and/or the finished exhibit. Interviewees generally included the director of the museum or other senior staff person responsible for the exhibit development, members of the exhibit design staff, fabricators, educators, and, in a few cases, evaluators.

We stressed to the museum staff that we were not there to evaluate the exhibits they were building, but that we wanted to learn, from their experiences in the project, about the nature of collaboratives and how the overall initiative was working. We found that by adopting a broader view of our mission, removing the primary focus from the judgment of their project, and setting a tone of cooperative learning, we were able to have frank discussions about their experiences with the ERC to date.

We designed the investigation around the model of an initiative discussed in Volume 1. We formulated general questions to investigate each link in the chain of events from proposal to final outcomes:

- What was the genesis of the collaborative? How did the collaborative form and get NSF funding?

- What is the character of the museum and its community setting? How has this character and setting influenced participation in the ERC?
- Are the project's goals congruent with NSF's goals?
- What is the collaborative actually doing? What kinds of collaboration exist? What kinds might exist?
- How does the exhibit design process work? The formative evaluation component? The collaborative activities?
- What are the staff members' perceptions of the quality of their and others' exhibits?
- What kinds of "outcomes" could the collaborative foster?
- What is the effect of NSF funds on the institution's ability to raise other funds? How does involvement in the ERC influence staff capacity and subsequent staff activities?
- How many visitors will actually see the exhibit? Is there any evidence that ERC exhibits have impact on visitors (attitude shifts, etc.)?

After each site visit, we wrote short (5- to 10-page) case reports. Those reports specified tentative hypotheses ("findings") about how the collaborative as a whole was working. For example, following one site visit we hypothesized that in the formative evaluation process, design staff tend not to adopt the formal, quantitatively based method of evaluation advocated by the technical advisor, but rather adapt their traditional, intuitive evaluative techniques to include more direct input from visitors. We then "tested" newly formed findings like this at each of the following sites we visited to gauge their validity for the collaborative as a whole. In this way, the information collected at each site was reviewed, and over the course of the study we were able to build through an iterative process a set of hypotheses about the essential features of the collaborative. Again, these hypotheses were not summary judgments on each museum's activities; rather, they provided a picture of the consortium's progress as a whole, to help NSF and museum staff refine and improve future activities.

Findings

Our findings concerning the Exhibit Research Collaborative are limited by our relatively short time at each site, by our emphasis on modeling a potentially effective procedure, and by the fact that the project is in the middle of a multi-year cycle. We cannot, then, draw conclusions about the success or failure of the participating centers' efforts. We can, however, describe the progress of the collaborative to date, and we can point out some important issues that need to be understood better. Below, we present a number of our findings to illustrate the results of the limited-case-study mechanism.

The ERC collaborative is functioning roughly as planned. After some initial scheduling difficulties, the participating centers are maintaining a realistic timetable for the development and circulation of exhibits. The first exhibits began traveling this past summer; the last exhibits to be built will end their travels in March of 1991. The current status of the collaborative is:

- Initial design phase: Charlotte, Boston, Lansing
- Prototypes completed/final design phase: San Diego, Seattle, Portland
- Exhibit built and circulating: Richmond, Louisville
- Received traveling exhibit: Portland.

By the fall of 1988, all museums will have built and begun circulating their exhibits.

Collaboration is a process requiring trust that must be learned. Like others that preceded it, this collaborative was set up to allow a large degree of autonomy and independence for the institutions involved. During the first phase of the project, the participating science centers shared their exhibit topic ideas at a series of meetings. The resulting self-criticism led to fundamental changes in two museums' exhibit topics and a refinement of the exhibit topic in a third science center. Interaction on such substantive issues and the resulting influence on the fundamental design concepts are rare among science centers or any other type of museum. Although we heard a handful of complaints about the harsh criticism of some exhibit topics, everyone agreed that the process of review and the resulting changes improved the conceptual foundation of the exhibits.

Since the exhibit topics were agreed on, the interaction among collaborative members has tended to focus on the technical details of building, shipping, and setting up complicated exhibits in diverse physical spaces. For example, fabrication staff from Richmond traveled to Lansing to help set up the "Night Visions" exhibit. Staff from Portland worked with individuals from Louisville over a long period of time to redesign and build new shipping containers for the "Systematica" exhibit. It is understandable that the collaborative has focused on the technical and logistical details of sharing the exhibits, but it is also clear that for a variety of reasons there has been much less interaction about pedagogical or substantive issues among mid-level staff (design and education staff, for example). Participants are well aware of this shortcoming and, with increased faith in the collaborative process, have begun to take concrete steps to address it. For example, in January 1988, the collaborative sponsored a meeting for exhibit design and education staff to facilitate cooperative efforts among mid-level staff. Some institutions even expressed a

13

desire to find ways to share the collective wisdom of the institutions by providing reviews of each other's exhibits, as well as suggestions for improving them. This kind of sharing obviously requires a climate of trust that appears to be building naturally, if slowly, through the experience of working with each other.

Many factors affect how each member institution builds exhibits and participates in the consortium. One goal of the ERC was to encourage participating institutions to design exhibits more purposefully and reflectively, with good data about the interests, motivations, and knowledge of the target audience. The following factors affected the degree to which each institution approached this goal:

- *Size of the museum in relation to ERC exhibits.* The match between the scale of the museum and the scale of the exhibit (1,500 square feet) played a large role in determining the priority given to the exhibit and the development process. For the relatively small, growing museums (San Diego and Louisville), which are struggling to establish themselves as mature institutions in their respective communities and in the larger science museum community, receipt of the NSF grant increased their institutions' credibility in the eyes of potential funders. For both, the ERC exhibit constituted one of the year's largest and most important projects.

  For the moderate-size, well-established museums (e.g., Richmond), 1,500-square-foot exhibits are about the right size, and the design and building of the exhibit has fit well into the standard operating procedures of each institution. For the larger museums (Seattle and Boston), participation in the ERC has become a secondary concern. In such settings, ERC exhibits have not added significantly to the museum's array of exhibits.*

- *Timing of the project.* The ERC project must fit into the overall schedule and exhibit development plans of each member institution. In at least one case, institutional schedules caused the ERC project to be built much faster than had been originally planned. It was clear that other exhibits and development activities can easily override the best-laid ERC plans.

---

* Boston has never been a midsize museum. The building of a 1,500-square-foot exhibit in the midst of projects four and five times the size is relatively low on the institution's priority list. The probability of significant institutional impacts is inversely proportional to the size of the institution involved; and, of all the centers, Boston's Museum of Science is probably the least likely to benefit or change from its participation. Similarly, the ERC exhibit has ceased to be a primary concern of the staff at the Pacific Science Center in Seattle. During the past year, this institution has celebrated its 25th anniversary and hosted the Association of Science and Technology Centers' annual conference. Within this context, the relatively small ERC exhibit has failed to capture as much attention as some of the larger exhibits designed and built over the past year and a half. A director of another, smaller museum did note, however, that larger, well-established science centers can contribute to a consortium by sharing their systems and experience with the smaller institutions.

- *Staff changes.* At several institutions, staff changes brought in new designers and administrators. These changes obviously affect the way in which the exhibits get built and shared.

- *Philosophy of the institution.* The design and educational philosophy of some of the ERC members was highly compatible with the proposed front-end design processes. Other institutions have different ideas about design (for example, they do not like unfinished prototypes on the floor) that caused a different approach to the design and building of each exhibit.

- *Institutional goals and political motivations.* Political and institutional motivations for participating in the ERC vary among the collaborative members. For Louisville and San Diego, the exhibit was a chance to build and demonstrate institutional capability to both internal and external audiences. The exhibit design and building staff at the Oregon Museum of Science and Industry (Portland) decided to turn their ERC exhibit into their annual "blockbuster." Consequently, the museum invested an additional $100,000 over and above the ERC requirements. The extra funds allowed for the inclusion of a large volcano in the "Nature's Fury" exhibit, as well as for a significant amount of publicity.

Because of these factors, a collaborative of institutions must expect and allow for variation in what each member will do. Uniformity is not possible or perhaps even desirable. The collaborative can emphasize certain shared features (e.g., in this case, a heavy emphasis on the design process), but it must allow members to adopt and express those features differently.

The quality of the science presented in the ERC exhibits appears to be high. Our 1-day visits to the participating museums did not permit us to carry out in-depth evaluations of overall exhibit quality. Moreover, six of the eight exhibits have not yet been finished. More important, it is not easy to get agreement on what is and is not a high-quality exhibit. We did, however, ask design staff about their perceptions of the quality of their exhibit (whatever stage of production it was in). The answers were uniform: design staff thought that their ERC exhibits were at least as good as or better than other exhibits that they had worked on. We encountered several cases where designers felt that the ERC exhibit was not particularly different from other exhibits, except perhaps in the emphasis given to the basic science presented. As one designer told us:

> When you have a corporate sponsor, you have to build the exhibit to include their interests (e.g., nuclear power, electricity, heart associations). When NSF is the sponsor, it is the same except that they are concerned about the quality and accuracy of the science presented.

Most of the designers were happy to have the "luxury" of focusing on science and of having the time and resources to devote to the design phase. One institution held a conference of subject matter specialists to brainstorm ideas for the exhibits.

Another museum used the ERC project as an opportunity to build an exhibit it had wanted for a long time that was too "sciency" for corporate sponsors.

There have been a few technical problems with the two exhibits that are finished and traveling. Richmond's "Night Visions" has had a difficult time standing up to the physical handling of the visitors at the Boston Museum of Science. When we visited Boston, the exhibit had been removed from the floor for repairs. "Systematica," built at the Louisville Museum of History and Science, also created problems for the staffs at Seattle and Portland. Staff in Seattle had a difficult time setting up the complex exhibit; and fabricators in Portland had to rebuild the shipping crates. It may be important that future collaboratives build in funds for a person to travel and help install an exhibit (for at least the first few sites). It also appears that exhibits can be further improved as they travel if there is a mechanism that allows institutions to suggest and make revisions.

Formative evaluation has been used in almost all sites and has helped the design of exhibits to evolve and become more successfully matched to the audiences that will use them. In theory, the formative evaluation process includes a series of stages in which the exhibits' design and content are assessed iteratively through visitor pretesting and the testing of successive prototype exhibits. There is no single formula for incorporating effective formative evaluation in exhibit design, but the process might include the following steps:

(1) Formulation of clear goals and operationally defined objectives by the design team.

(2) Pretesting of visitors' knowledge of, interest in, and attitudes about the proposed exhibit topic.

(3) Reformulation of the goals and objectives based on the visitor pretest.

(4) Design and construction of "soft" prototypes.

(5) Floor testing of the soft prototypes, by assessing visitors' reaction to them and ability to manipulate them, as well as the visitors' ability to comprehend the science content of the exhibit.

(6) The redesign of the exhibit based on the floor testing.

(7) The building of "hard" prototypes of each of the exhibit components.

(8) The floor testing of the hard prototypes.

(9) Final design changes based on testing of full prototypes.

(10) Fabrication of final exhibit.

To assist staff at the participating centers in carrying out the formative evaluation component, the consortium provided for a professional evaluator from Richmond to spend a minimum of 1 week in each museum. The evaluator visited each of the centers at the beginning of its design process. She helped to clarify goals and specify objectives, helped staff to design visitor survey instruments, practiced with staff as they interviewed visitors, and assisted in the development of a plan for subsequent evaluation activities. In addition, the evaluator remains available to science center staff via the telephone, mail, and conferences to provide further technical assistance.

Our site visits have shown that the assistance the evaluator provides is crucial to the science center staff as they try to grapple with the relatively unfamiliar tasks of formative assessment. At each museum, at least one staff member noted that it would have been impossible to carry out the formative evaluation without on-site technical assistance. With the evaluator's help, each site has undertaken relatively formalized evaluation activities. In all but one case, this assessment component represents a sharp break with previous practice. Few staff had ever participated in structured assessments of their exhibit designs at an early stage of the design process, and no center except one we visited had instituted regular mechanisms for evaluating the exhibit design process. The single exception is Richmond, where in-depth formative evaluation has been a part of standard design procedures for a number of years.

Given similar training and money, however, the centers still took very different tacks in carrying out their formative evaluations. In some cases, differences among museums can be traced to specific constraints facing the institutions. For example, Louisville faced an extremely short time frame within which to design and build its "Systematica" exhibit. Consequently, its formative evaluation process was less thorough than at some of the other institutions where staff had ample time to formulate and implement an assessment process. Institutions also varied a great deal in the expertise of staff in evaluation activities. Both San Diego and Richmond, for example, have staff with extensive evaluation experience, whereas Seattle and Louisville do not. Thus, it was easier for Richmond and San Diego to implement a full evaluation program.

Differences among the science centers' implementation of the formative evaluation component also reflected their varying philosophies of what constitutes an appropriate and effective assessment strategy. In particular, museums placed different degrees of importance on the objectivity and empirical grounding of the evaluation. Those centers that adhered to a more objective, empirically based assessment strategy tended to rely more on staff outside of the immediate design team to perform the evaluation activities. For example, Seattle hired an expert evaluator to work exclusively on the ERC exhibit design process. Similarly, Portland used in-house staff with evaluation experience who had nothing to do with the design of the exhibit. As one staff member noted, "We were afraid that members of the design team would be blind to flaws in their own work. It seemed like the only way to assure objectivity was to use 'outside' evaluators."

In contrast, those centers in which staff expressed less concern about objectivity involved design staff more directly. For example, at San Diego and Richmond, the design team was regularly involved (along with professional evaluators) in the formal evaluation activities at different stages of the design process. At these institutions, the concern was less with objectivity and more with the need for designers and educators to interact directly with visitors to gain first-hand knowledge of the audience's needs and interests.

At each of the institutions in which an exhibit has reached the prototype stage, the process of up-front assessment has led to basic changes in the design and fabrication of the exhibit. We saw several examples of how hands-on testing with visitors can help staff design and build an exhibit that can be better used by its intended audience without sacrificing the original science content. Our case visits also demonstrate how difficult the process is. In one case, a full year elapsed between the time staff began tentative designs for the exhibit and the time it went into final fabrication. Across all the institutions we visited, the effects of the formative evaluation process ranged from minor technical changes to key reformulations of the design concepts.

We also saw that where the formative evaluation process was too formalized, it conflicted with the informal, intuitive assessment techniques implicitly used by exhibit designers. Exhibit designers have long "evaluated" their exhibits. That is, they have thought through their ideas, discussed issues with colleagues, reflected on past experience, drawn up tentative drawings, and rethought ambiguous ideas. This "connoisseurship" is intuitive and aesthetic; it often stands in sharp contrast to what is perceived to be the objective, quantitative world of conventional formal evaluation (museum staff are typically unaware of developing evaluation traditions that emphasize qualitative data and connoisseurship). In the world of designers, the data are subjective and qualitative; the assessment "instrument" is the designer him- or herself. In the world of evaluation (as perceived by these museum staff), data are typically quantitative, and instruments are seen as objective tools used to measure accurately the phenomenon under study. Our site visits found staff grappling with the contradictions between the two approaches.

Our case studies suggest that the built-in support for front-end assessment procedures provided exhibit staff with the impetus and opportunity to spend more time thinking through the exhibit, justifying it to others, and, most importantly, interacting with visitors. That is, the existence of a formative evaluation component that each center was supposed to go through encouraged exhibit staff to spend more time in an interactive manner, reformulating the design of the exhibit--an opportunity that their normal operating procedures, money, and time constraints usually preclude.

We did not find that exhibit staff became expert evaluators. In fact, they most frequently made decisions on impressionistic evidence (as they always had done), but this evidence was gathered through greater and more intensive interaction with other staff and visitors. For example, we found that the process of standing in front of a prototype for a few hours and watching visitors' behavior or asking a few questions "told" the designers most of what they needed to know. They rarely had to wait for tabulations of the amount of time visitors spent at each exhibit or the average number of correct responses to a questionnaire. The formative evaluation component may have "worked" simply because it pushed staff to spend more time thinking through their designs from the point of view of the visitor. In fact, in some centers, too much emphasis on proper technique and quantitative data may have hindered the design process. For these reasons, it seems important that institutions not completely separate the design and evaluation process by hiring full-time evaluators who carry out extensive studies and then report their results to the designers.

ERC members "believe in" formative evaluation but face formidable barriers in institutionalizing the process. One of the central goals of the founders of the ERC was to help participating institutions develop an expertise in, and commitment to, formative evaluation. That is, one of the most ambitious and long-term goals of the collaborative was to change the modus operandi of exhibit building so that front-end assessment became a built-in, accepted part of the exhibit-building process.

How fully will this goal be realized? One institution pretty much operated this way to begin with. For at least one of the smaller institutions, this goal seems to have been quite fully realized. For the larger institutions, our findings suggest that the impact will probably be focused on a few designers and not influence institutional procedures to any great degree.

Thus, although this study shows that the staff at participating centers have had a significant and positive experience with formative evaluation, the science centers still face significant obstacles in building such techniques into their regular design process. The first and most formidable barrier is the significant amount of time (and money) that it takes to carry out meaningful formative assessment. This requires institutions to shift scarce resources from the back-end fabrication of exhibits to the front-end design process. All but two of the centers said that it would be unrealistic to expect to allocate the same percentage of resources to future design processes as they had for the ERC exhibit. (There is also an irony here that some institutions came to think that evaluation could be done only in a rigorous and large-scale fashion and thus was an all-or-nothing proposition.)

A second major barrier is staff expertise. As the staff at each center look to the future, they invariably raise the question of who will have the expertise to carry out or even supervise evaluations in subsequent projects. They realize that a professional evaluator will probably not be there on the next exhibit. Most staff believe that, although they are much more comfortable with evaluation in general, they do not have the skills to perform one on their own. Consequently, centers are weighing various options for the future. An educator in Louisville is working overtime and using volunteers from the Junior League to carry out pretests on new exhibits, while the director is trying to raise funds for a full-time evaluator. In Portland, one staff person has volunteered to take classes in evaluation design and become the in-house resource person. San Diego has sufficient staff, but not necessarily sufficient time. Staff in Seattle doubt that they will ever be able to recreate the evaluation component of the ERC design process.

The ERC traveling exhibits do not reach more people than do permanent exhibits, and their cost per visitor is roughly the same. In other words, investing in the ERC consortium is not more cost-effective (in visitor interaction-hours per dollar) than are investments in permanent exhibits at one institution. Because of the relatively small size of the ERC exhibits, they do not function as a draw for visitors, and museums generally do not find it worthwhile to run a publicity campaign for them (with the exception of Portland, which had invested significant extra resources in its exhibit). In general, then, we can expect that the exhibits will be seen only by the regular visitors at each museum. Although this number of visitors will be substantial, each science center would have to attract more than 150,000 visitors a month--more than three times the expected average monthly visitors--to reach the projected 40 million total (exhibits will be in centers for a total of 258 months during the lifetime of the project).* After 1991, when all exhibits have finished traveling to participating centers, some centers plan to lease them. This step would increase the total viewership.

The investment in the ERC collaborative appears to achieve significant leverage in several areas. The leverage gained by NSF's investment comes in several forms. One conspicuous form of leverage is the effect of "peer pressure." The collaborative, in effect, serves as a second-level review process and standard-setting body. Because they were to share their exhibits, many of the institutions worked hard (and invested more of their own funds in the exhibits) to make sure that their exhibits would be judged well by the others. Also, to a limited extent, collaborative members helped each other to think about their exhibits.

Another leveraging function of a collaborative is to share resources. The evaluator who helped each of the museums is an example of a scarce resource that can be efficiently shared by all of the participants. (Also, the evaluator served an important function as a "link" between all of the institutions, informing each about the others' activities.)

Probably most important, the collaborative mechanism allows NSF to invest in medium-size museums on the same basis that it funds exhibit development in the larger museums. Because the investment not only results in exhibits being produced but also indirectly builds the capacity of these institutions to do further exhibit building, NSF can use the collaborative mechanism to further the overall development of science museums across the nation. Without such mechanisms, NSF might find itself able to fund only the largest institutions, thus helping the "rich get richer." Also, by increasing the stock of good traveling exhibits, NSF may indirectly be helping small science museums that do not have the capability to build their own.

---

* The 40 million figure may represent "proposal hyperbole." The figure was calculated by combining audiences at institutions to which the exhibits were sent plus the home audience that would be able to view a permanent prototype at the originating institution.

Finally, NSF's prestige and credibility significantly increase the ability of small museums to attract other funders, as well as boosting their stature in the local community. In addition, the fact that the exhibits will circulate nationally makes them potentially more attractive to other funders who want to invest in science education improvements at a national level. The consortium might serve as a useful "umbrella" for the member institutions and allow a fund-raising approach that no single institution could carry out alone; the collaborative has already served this purpose by making the ERC proposal attractive to NSF.

Implications of Findings for NSF

Our limited case study suggests that the collaborative mechanism is an effective way to fund museum exhibit development, particularly for the midsize, maturing science museum. The collaborative appears to be producing high-quality, interactive science exhibits that equal or exceed the best work done in the participating centers. Importantly, the ERC is successfully getting good exhibits out to a large audience (although probably smaller than originally projected) in a cost-efficient manner. The ERC has also demonstrated the power of the collaborative to place an emphasis on the exhibit design process, fostering meaningful and effective formative evaluation, even in instances in which staff have had virtually no evaluation experience. Finally, the ERC represents a mechanism for NSF to enhance the professional capacity of the field of exhibit development.

Within these general findings, we find great variation among the science centers in their approaches to designing and assessing their exhibits, in the extent to which these processes differ from their standard operating procedures, and in the degree to which they will be institutionalized in the centers. And we expect there to be important variations in the quality of the exhibits, although it is too soon to make meaningful judgments on exhibit quality.

Naturally, the success (as well as the limitations) of the collaborative is primarily a function of the institutions' existing professional capacity and commitment to high-quality, educationally effective science exhibits. But our site visits also suggested that the technical and administrative support provided by the collaborative fosters more effective design processes. The fact that the grant included some funds for support activities made it possible for the project leaders to focus on guiding the project through its development. The existence of an evaluation expert to help museum staff design and begin formative assessment procedures was another crucial support. It seems reasonable to conclude that a number of the institutions would have been hard pressed to mount evaluation efforts without outside technical assistance.

In fact, it appears that more structure and technical assistance might strengthen the collaborative's efforts. In a number of the centers, a 1-week visit from the evaluation expert was probably insufficient. A full-time technical assistance professional would have helped the exhibit staff to create more coherent evaluation designs. Such assistance appears crucial in projects designed to foster new and complex behaviors among professionals, as was the case with the ERC. Similarly, the collaborative effort might have been helped by a more clearly developed strategy for fostering collaboration among staff at all levels of the design process. For example, future projects might want to plan concrete activities for educators and designers to share their knowledge and to assist one another in tackling tough issues in the design process. Ultimately, collaboratives might want to have a structure and process for reviewing the results of their work and helping each other learn from their experience. The ERC has had to struggle with these issues on an *ad hoc* basis; reviewing each other's work took place only after participants had dealt with the myriad technical details that arise when eight institutions build exhibits for one another.

We would suggest that future projects of this sort plan and allocate resources for a range of collaborative activities in their proposals. NSF might encourage this feature during the proposal solicitation and review process; the Foundation might also consider funding technical assistance contracts in conjunction with collaborative projects.* The technical assistance contractor could hold responsibility for administrative details (which at times threatened to overwhelm the ERC leadership), convening meetings, and helping staff throughout the process to incorporate formative evaluation techniques.

Our site visits also raise an issue about the homogeneity of consortium members. One can never hope to find 8 or 10 centers with exactly the same needs, institutional capacities, and philosophical approaches, but encouraging some comparability appears essential. It is clear that Boston is too large and well established an institution for this collaborative aimed at midsize museums. Although San Diego handled its participation well, its space limitations may have put too many constraints on the other centers' exhibits. There is also an issue of philosophical compatibility--would the consortium work better if it were composed of "like-minded" institutions or does diversity help institutions learn from each other? Should these same institutions continue to interact in a second consortium project, or should they split into two groups of four institutions, with each group taking on four new institutions? We simply note that potential consortium members, and possibly NSF, should pay close attention to the types of institutions participating in future collaboratives.

Finally, our study points to the need for more research on, and experiments in, building institutional and personal capacity at the same time that exhibits are

---

* The Ford Foundation successfully uses outside contractors to provide evaluation as well as technical assistance to the consortia it funds.

27

developed. The collaborative is a mechanism that can influence how institutions "do business" (for example, by encouraging and building the capacity to do meaningful formative evaluation in the exhibit design process). With guidance, the collaborative might lead to stronger networks and more shared work among science museums, which in turn would help meet an important need for professional development. With assistance, the collaborative might also become more sophisticated about using the leverage of this mechanism in its fund-raising efforts. All of these questions of capacity-building rest on the premise (borne out by our case study) that the collaborative mechanism can be used to promote the collective and individual interests of the participating museums simultaneously. Our findings suggest that the collaborative can be structured so that the benefits of autonomy are not lost, while the benefits of cooperation are maximized. Assessments of experiments like the ERC should help in learning ways to do this.

Whether our findings imply different courses of action for NSF depends on how purposeful and proactive a stance the Foundation chooses to adopt in this or related areas of investment. If the Foundation assumes a reactive posture in response to other field-initiated proposals for collaboratives, then our findings are primarily advice to proposers and to NSF as it reviews proposals and encourages refinements in them. If, on the other hand, NSF chooses to pursue this line of investment more proactively, it may wish to solicit collaborative proposals and provide support to them in a more purposeful way. Or it may wish to encourage more indirectly the formation of collaboratives and consortia. In that event, our findings suggest various ways that the Foundation can undertake and fund the collaborative exhibit development (and collaborative work in related areas of investment) as a formal initiative.

II DESCRIBING THE DOMAIN OF INVESTMENT THROUGH SYNTHESIS AND ANALYSIS OF SECONDARY DATA: A "MACRO VIEW" OF INFORMAL SCIENCE EDUCATION

By

Andrew A. Zucker

II DESCRIBING THE DOMAIN OF INVESTMENT THROUGH SYNTHESIS AND ANALYSIS OF SECONDARY DATA: A "MACRO VIEW" OF INFORMAL SCIENCE EDUCATION

This "macro view" of informal science education is an attempt to create a statistical profile of the field as a whole, focusing whenever possible on children up to the age of 18. It is limited by the availability of data, and also by time constraints. The result, therefore, is "broad brush" and incomplete. We have concentrated on television and museums because of their important place in NSF's strategy for supporting informal science education activities.

We are aware of only one study that has focused on the entire spectrum of informal science education activities for an individual or for a population: the joint Corporation for Public Broadcasting (CPB) and U.S. Department of Education Home Information Technology Study (CPB, 1986a; Riccobono, 1986). Generally, what knowledge we have of informal science education is available "piecemeal" (e.g., for one limited aspect of a particular medium), making it difficult to form an accurate impression of the way in which informal science education as a whole plays a role in people's lives. Indeed, a committee of the National Research Council has recently recommended additional data-gathering concerning out-of-school science and mathematics activities (Murnane and Raizen, 1988).

Highlights

Readers should approach this statistical profile as a feasibility demonstration rather than a finished research product. In addition to the limited scope of the review, there are complicated questions about the adequacy and interpretation of existing statistics that could not be resolved in the limited time available for this task. A more rigorous and complete macro view of informal science education would have included more attempts to cross-check and interpret "suspect" statistics and to employ external review of the analysis by expert consultants. For example, relying on industry figures alone for estimates of public-television viewership is weak, because publicly available figures from these sources serve many functions, including promoting public-sector television. A more thorough analysis would have enabled us to contrast estimates from different sources and adjust accordingly. To do a more rigorous review, of course, requires more resources; because the domain of informal science education encompasses many areas of NSF's investment, the effort to understand the domain may well justify the expense.

From available statistical data, we have created a sketch of informal science education in which the following themes emerged. First, regarding the public and its use of its time, we found that:

- Excluding time for work (or school) and sleep, Americans, on average, put at least a third of their total weekly time into activities (television, reading, crafts) that can involve informal science education.

- Approximately one-fifth of a national sample identify leisure-time pursuits with a significant scientific component as their most important informal learning activity (which they may or may not acknowledge as "scientific").

- Orders of magnitude can be assigned to the amounts of time Americans attend to different informal science media: for example, the public devotes an estimated 60 times more hours, on average, to public TV science viewing than to visiting a science museum.

Second, regarding television as a source of informal science education, we found that:

- American audiences watch 20 hours of commercial television, on average, for every 1 hour of public television.

- Approximately three-fifths of the public "attentive" to science policy, two-fifths of the "interested" public, and relatively few of the "noninterested" public regularly watch science shows on television.

- A very high proportion of children's programs on commercial television involve at least a theme or aspect explicitly and unambiguously related to science (including space or science fiction). Nonscience television programs such as dramatic series and the news convey a great deal of the public's information and misinformation about science.

- Programs about animals, science, and nature are a highly valued part of the public television schedule.

Third, regarding the availability and use of science museums, we found that:

- Science museums are increasing in number and are extremely popular; they are visited by numbers far out of proportion to their representation in the museum population.

- About half of science museum visitors are children. Data about the composition of the visitor population are extremely weak (and NSF is supporting survey work that will partially remedy this situation).

Method for Constructing the Macro View

The basic approach was to proceed from the bottom up--that is, to collect a great many factual items and data tables first, and then use these to construct a picture of the informal science education field. In practice, there was often an iterative process at work, in which the existence of certain data would help fill in the picture but would also underscore the existence of a "hole" that needed to be filled with data not yet gathered. For example, knowing the ratings of a number of science shows on public television led us to wonder about the ratings of commercial science shows, and we then proceeded to fill that particular "hole." Whenever possible, a general-purpose statistical reference, such as the *Statistical Abstract of the United States* (U.S. Dept. of Commerce, 1987) or *The Condition of Education* (U.S. Dept. of Education, 1986), makes an excellent starting point for research, precisely because such sources often focus on "the big picture."

Many different sources of data were tapped, beginning with the shelves and file cabinets of the researchers:

- A variety of libraries were used--for example, in the case of museums, the Smithsonian Museum Reference Center, reference material at the headquarters of the Association of Science-Technology Centers (ASTC), the George Washington University Gelman Library, and several of the Stanford University libraries. A majority of the libraries used are now catalogued on computer (or compact disc, in one case), making searches faster and easier than in the past.

- Discussions with experts in the field proved useful, both for preliminary research and to answer specific questions. Staff at the Smithsonian Museum Reference Center, for example, were able to provide access to numerous documents in response to our requests for a general orientation and for visitor surveys. The directors of research for the Corporation for Public Broadcasting and for the Public Broadcasting System were very helpful in answering our requests for certain specific data relating to public-television viewing.

- Numerous documents were obtained especially for this analysis from sources in several different states. For example, results of a Field Institute/California poll were provided at low cost after the nature of our research was explained (Field Institute, 1985). The Public Opinion Laboratory at Northern Illinois University provided us with a variety of useful reprints.

Informal Science Education in Context

"Informal science education" is a simple term that covers a very large and complex set of activities. In a nutshell, informal science education is the education in science that people (both children and adults) receive in nonschool, noncredit settings. For purposes of the current inquiry, science is meant in its broadest sense, including mathematics, engineering, and technology generally.

Informal science education takes place using whatever media and settings are available and that individuals choose to use. The "channels" for informal science education include television, print media (e.g., books, magazines, and newspapers), museums, zoos, aquaria, hobbies (e.g., gardening), doctors' visits, social conversations, and many others.

Before we sketch the nature and the dimensions of specific "channels" for informal science education, some information about the public's understanding of science and how people use their nonwork, nonschool time provides a context for understanding this investment domain. As with all sections of the "macro view," these short overviews are intended only to provide a brief orientation, not to be an exhaustive or scholarly review of available data.

Public Understanding of Science

Although it is generally agreed that the objective of education in science is to produce "scientific literacy," there is considerably less agreement as to what the term ought to represent, how scientific literacy can be measured, and what purposes scientific literacy ultimately serves (Miller, 1986, 1987a, 1987b; Laetsch, 1987). In general, however, there is widespread agreement that the scientific literacy of American students and of the public at large is too low, and that it should be increased (NSB, 1983).

The most recent published data from the National Assessment of Educational Progress (NAEP) showed an overall continuing decline in student achievement since 1969, when it was first measured by NAEP (Hueftle et al., 1983). The Second International Science Study showed that, 17 years later, the science achievement of American students still lags behind that of students in other countries--even for advanced science students (IEA, 1987). Fewer than half of American 13- and 17-year-olds think that their science classes are interesting (Hueftle et al., 1983); and once students at the secondary level are given a choice, most soon drop science, so that fewer than one-third of high school graduates have had 3 or more years of science (NCES, 1984).

A prominent series of studies of scientific literacy for the National Science Board defined the concept in terms of three dimensions: understanding the processes of science; understanding some basic terms and concepts (e.g., "radiation," "DNA"); and understanding the impact of science and technology on society. Based on a 1985 survey of over 2,000 adults, this research concluded that only 5% of American adults qualified as scientifically literate--a slight decline since 1979 (Miller, 1987a).* An interesting aspect of his data is that a much higher percentage of the public is "attentive" to science policy than is actually scientifically literate; specifically, whereas roughly 20% of the American public is attentive to science policy, only about

---

* Miller goes on to conclude that "the informal science education efforts of recent years have not produced any measurable increase in scientific literacy." Experts disagree on the soundness of this conclusion and the percentages assigned by this research to each category of the public. Miller's estimates are based on responses to a relatively small number of survey items.

10% of those attentive to science policy are judged scientifically literate by the standards he has established (Miller, 1987a).

Corresponding to relatively low levels of scientific literacy are relatively high levels of "scientific illiteracy." Most commentators believe "illiteracy" about science, mathematics, and technology is widespread, regardless of how it is measured. In general, there is a high correlation between a person's level of education and his or her level of scientific literacy.

The Public's Use of Its Time

How does the public use its time? Average figures for the population are presented here. Where possible, the discussion of particular "channels" has included breakouts of average figures, e.g., by age and sex. The data on Americans' use of their time are summarized in Figure II-1. These data, which have been synthesized from a variety of sources, should be considered only approximations.

Work (or school), sleep, and television account for nearly two-thirds of the average American's time. In 1985 the labor force of the United States was about 107 million, compared with a total population over age 16 of about 180 million. If both employed and unemployed persons are included, the average work week was about 21 hours (U.S. Dept. of Commerce, 1987). The average school week for children in school is on the order of 30 hours. These figures represent at least roughly people's total "work" hours outside the home. Sleep accounts for about 55 hours per week (U.S. Dept. of Commerce, 1980). An overall U.S. average of about 29 hours per week goes to television viewing (Traub, 1985).

A 1975-76 study of time use among American households provides additional data about the remaining one-third of people's time (Juster, 1986). In order by the amount of time per week spent on each activity, the most important activities were: socializing, cooking, shopping, reading, cleaning, sports, organizations (including church), child care, crafts, repairs, and spectator events (including movies). Presumably, what time remains unaccounted for in these categories is taken up primarily by commuting, eating, and personal care.

Americans' reading time was about equally split between newspapers on the one hand, and books, magazines, and the like on the other (U.S. Dept. of Commerce, 1980). However, the averages mask significant individual differences. One survey found that in 1983, 50% of respondents had read one or more books within the past 6 months; 44% had read, but not books; and 6% were nonreaders (U.S. Dept. of Commerce, 1987). The 1985 Home Information Technology Study found that books and magazines were the most frequently used and trusted "learning resources" by all age groups (Riccobono, 1986).

Many common activities have some component of science involved, particularly if natural history is included. For example, in 1980 fully 25% of the U.S. population

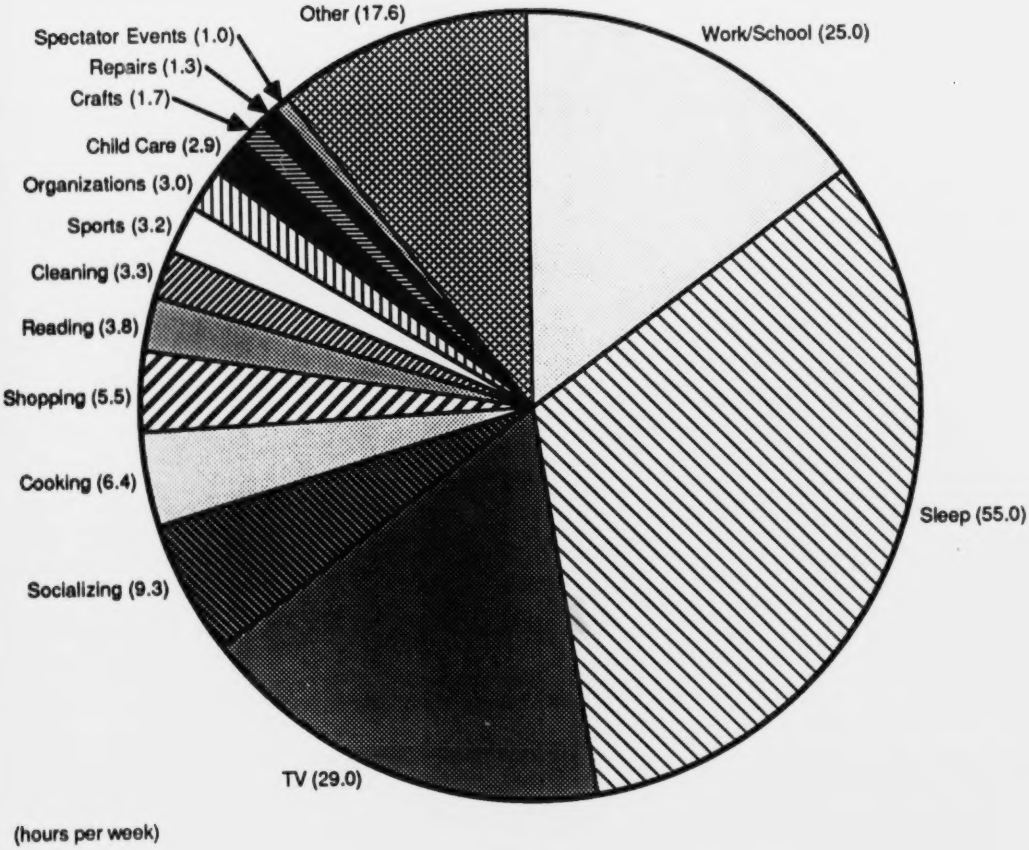


FIGURE II-1 AMERICANS' USE OF THEIR TIME

over age 12 were characterized as "sportsmen," meaning that they engaged in hunting or fishing. Similarly, in 1985, 37% of all households were reported to have vegetable gardens. According to a 1985 Gallup poll, the five most popular leisure activities are flower gardening, swimming, fishing, bicycling, and bowling (U.S. Dept. of Commerce, 1987).

If people are asked what informal learning activities they have engaged in, learning science or mathematics is considered "most important" by only 4% of the respondents (CPB, 1986a). However, if learning activities focused on computers, animals/nature, health/hygiene/safety, and first aid/lifesaving are added to science and mathematics, then 19% of the respondents found these to be their most important informal learning activities. Other categories used in reporting the survey results (such as crafts, camping, and civics/ government) may also include science-related activities (CPB, 1986a).

Statistics focusing on aggregate uses of people's time seem both helpful and frustrating. Reminders of the importance of social contacts (e.g., a teenager discussing auto engines with a friend) and of print media, such as newspapers, are helpful, since these are examples of aspects of science education that probably receive too little attention in most discussions of the subject.

Aggregate statistics also make it possible to assign orders of magnitude to the public's exposure to science information via different channels or media. For example, data collected for this macro view enable us to make some rough estimates of the per capita time spent by the public on informal science activities at science/ technology centers and museums, as well as on the time spent viewing science on public television. The results are only estimates, but at least they help in determining the order of magnitude of the time spent on these activities.

ASTC reported that its member museums had 38 million visitors in 1979 (ASTC, 1980). On the other hand, the IMS/NCES survey of museums in the early 1970s indicated visits to science museums at about 117 million--a very different figure! Much of the difference is presumably due to differences in the universe of science museums covered by each survey. We might choose an intermediate figure of 50 million for purposes of comparing orders of magnitude. The museum calculation then goes:

> *Science museum viewing*: Total annual science museum visits are 50 million. Multiply by time per visit of 1.1 hour, derived from an old Smithsonian visitor survey (and a very generous estimate in light of others' assertions that the visitor spends only about 30 minutes attending to exhibits--see Nichols et al., 1984). Divide by 52 weeks per year. Divide by U.S. population over age 5, which was 221 million in 1985. The result is about 0.005 hours per week per person of science viewing. That is about 0.3 minutes (*17 seconds*) per person per week, or about 15 minutes per person per year.*

---

* This figure, in effect, averages those who do and do not go to museums. It does not provide a realistic indication of the total average exposure to science among museum goers.

Looking only at science on public television, a very rough calculation then goes:

> *Public TV science viewing*: Average TV viewing time per person is 29 hours per week. Multiply by 5% for public TV's proportion of the whole. Multiply by about 20% for the proportion of public TV devoted to science (a very rough estimate based on "Public Television Programming Content by Category, Fiscal Year 1982," published by CPB). The result is about 0.3 hours per person of public TV science viewing per week, or *15 hours per person per year.**

At the same time, a really complete list of important informal activities is too long and too detailed to show up in gross figures such as those cited. Museum visits, for example, do not appear, because they are a tiny fraction of the average person's week. A few other examples of potentially important informal science activities showing small amounts of time per capita include: participation in clubs with some science component (e.g., 4H, Girl Scouts); part-time and summer jobs for youths; intentional yet informal education at doctors' and dentists' offices (e.g., film loops); the use of science-related toys (microscopes, gyroscopes, etc.); and, the growing use of computers and VCRs for home-based education (see Section III). For NSF, peoples' use of their time is important, yet it is only one variable that needs to be considered when investing in informal science education.

Television and Informal Science Education

In 1986, about 98% of all U.S. households contained a television set. More than half contained two or more sets, raising the average number of sets to about 1.8 per household. In the average household, a television set is on more than 7 hours each day (U.S. Dept. of Commerce, 1987).

There are more than 1,230 television stations in the United States. Of these, about three-fourths are commercial stations, while the remainder are public, i.e., nonprofit. About 40% of American households are now served by more than 7,000 cable television systems (U.S. Dept. of Commerce, 1987). In 1980, new programming for all television sources was produced by only about 20 television producers (U.S. Dept. of Health and Human Services, 1982).

On an average weekday evening during "prime time" (e.g., 8:30 to 9:00 p.m.), about 105 million people are watching television. This is nearly half the total population of persons over age 5 (Traub, 1985).

---

* See footnote on previous page. This figure represents an average across the entire population, not the total exposure to science television among those who opt to watch science shows on public television.

The Television Audience

For all persons in the United States ages 2 and above, the average amount of TV viewing is nearly 29 hours per week, or over 4 hours per day. Figure II-2 shows the amount of television viewing by various portions of the population, as of October 1984 (Traub, 1985).

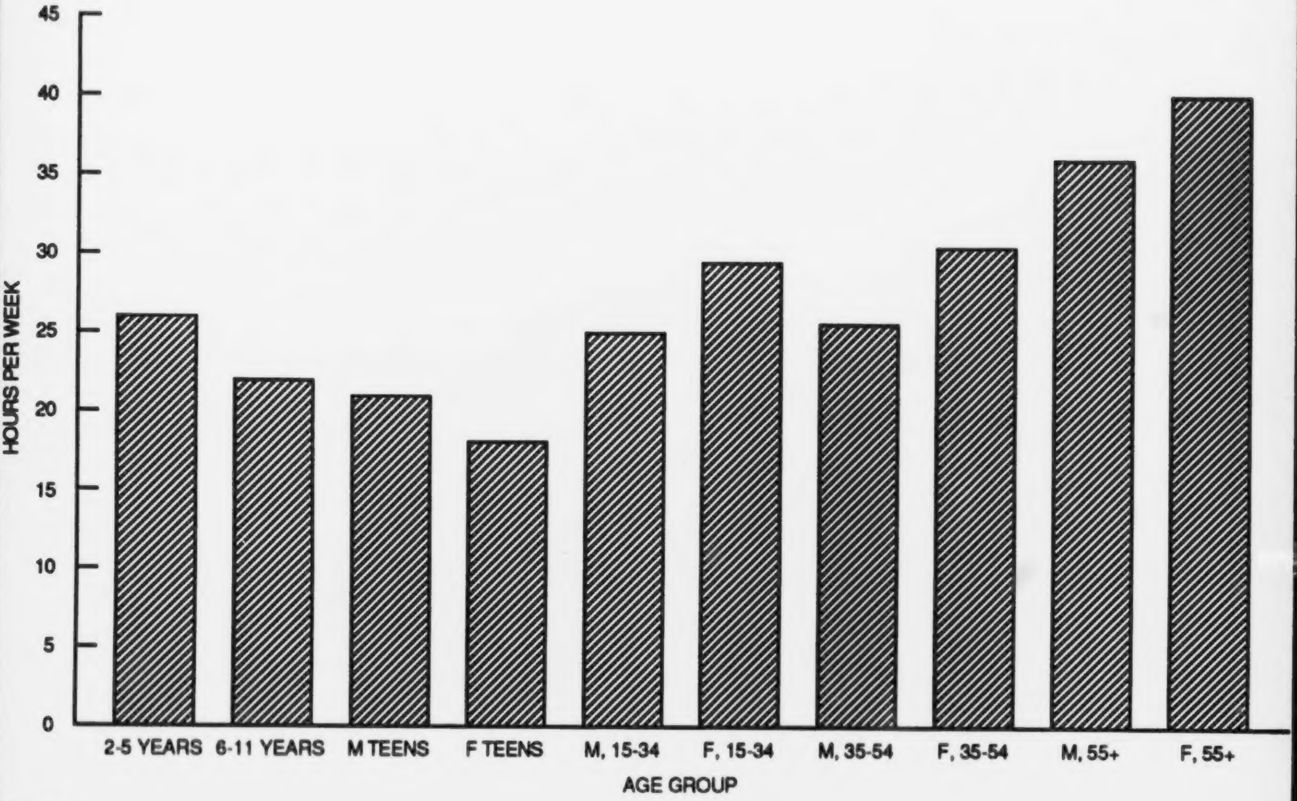
- The heaviest TV viewers are women ages 18+ and men ages 55+ (Traub, 1985).
- Children ages 2 through 17 are actually below-average TV viewers. Nonetheless, 2- to 5-year-olds watch nearly 26 hours per week, and 6- to 11-year-olds watch nearly 22 hours per week (Traub, 1985). Most children, most of the time, watch adult television (Tressel, 1984).
- Male teens watch about 21 hours per week and female teens watch about 18 hours per week (Traub, 1985). ("If your target is teens," says Barry Kaplan, a vice president at the Ted Bates advertising firm, "network television is a waste.")

Studies tend to show that children's leisure-time television viewing has an impact on achievement in school. The effects are slightly positive for up to 10 hours of viewing a week, but beyond this the effects are negative and increasingly more deleterious. Females and high-IQ children are more adversely affected than other groups (Williams et al., 1982).

About 95% of the audience's viewing time is spent watching commercial television, and only 5% is spent watching public television (U.S. Dept. of Health and Human Services, 1982; PBS, personal communication). In an average week, however, nearly 80% of television households tune in to a PBS station once or more often (CPB, 1986b).

TV is a mass medium, which depends on attracting a mass audience. For example, in a recent week the top-rated network prime-time show ("The Cosby Show") was viewed by more than one-third of all TV households. Even the lowest-ranked of the top 20 shows was viewed by more than one TV household in six, or over 15 million households (*Washington Post*, 1987). An audience of "only" several million people is very small by commercial television standards, and simply does not have much "clout" in the industry.

As might be expected, American adults who have a high level of interest, functional knowledge, and a pattern of relevant information gathering (24% of the adult population by one estimate) appear most likely to watch science shows on television. Data from a 1981 survey show that 59% of this "attentive public" were regular viewers of a science television show, compared with 40% of the adults whose interest was high but who lacked knowledge of science (another 28%). "National Geographic Specials" were the most frequently viewed by both groups (Miller, 1983). Although we have no data on the science viewing habits of the half of the public (roughly) who were not



M = Male
F = Female

FIGURE II-2 TELEVISION VIEWING: WHO WATCHES HOW MUCH

attentive to science policy, it seems reasonable to speculate that many fewer than 40% were regular viewers of science television shows.

Science on Commercial Television

A very high proportion of children's programs on commercial TV (60% to 70%) involve at least a theme or aspect explicitly and unambiguously related to science, including space or science fiction. Similarly, science and technology themes appear in about half of all dramatic network programs. Because regular science news and science programs attract relatively small audiences (7% to 10%) in comparison with these general and dramatic network programs, the latter are probably a more important source of information affecting viewers' attitudes, information, and opinions toward science (Gerbner et al., 1981).

One of the most successful science shows on commercial television is "Wild Kingdom," which went on the air in 1963. During the 1982-83 season it aired on about 300 stations. In July 1987, "Wild Kingdom" had Arbitron ratings of about 2, with an audience share of 9. A well-known example of a failed commercial science show is Walter Cronkite's "Universe," a "magazine format" science program that originated in 1980. Despite CBS' hopes of attracting a large audience for "Universe," the program was dropped in 1982 because of "low" ratings--an audience share under 20 (Tressel, 1984).

Although there is relatively little evidence to go on, some research has examined the image of scientists and science that commercial television projects. Scientists comprise fewer than 1% of prime-time characters portrayed in the work force--half the actual percentage. Health professionals, including doctors, are portrayed at seven times their rate in the actual population. Although the scientists' aggregate personality profile was judged to be positive by researchers, in comparison with health professionals and other prime-time characters scientists are shown as relatively less attractive, fair, sociable, warm, tall, young, and peaceful. Depicted on weekend children's programs, they were also judged to be less rational and stable, and much more violent, than other characters (Gerbner et al., 1981).*

Relatively little can be said about the quality of scientific information conveyed by commercial television, but it is reasonable to suppose that the information is fragmentary and often inaccurate. A study of prime-time shows conducted in the late 1970s, for example, showed that more health information is provided about infrequent and unfamiliar health problems than about common and widespread ones, and that little specific health information is provided. A similar study in the early 1970s found that only 30% of the health information that was provided was rated as useful, whereas 70% was considered inaccurate, misleading, or both (U.S. Dept. of Health and Human Services, 1982).

---

* The characterizations we cite represent the conclusions of one team of researchers. As yet there have not been other studies that confirm or refute these patterns.

Science on Public Television

By comparison with the commercial networks, public television audiences are generally much smaller. For example, "The Sharks," a PBS National Geographic special with the highest Nielsen rating in the history of public television, was viewed by 17.4% of TV households, which is just slightly more households than viewed the 20th-ranked network prime-time show in a recent week. The average rating for "Nova" in 1985 was 7.8%; for "3-2-1 Contact!" it was 4.3%. Many PBS offerings average a rating of 2 or less (Chen, 1984).

Because there is often more than one viewer per household, and because viewers may watch only occasional shows in a regular series, the total audience for a series may be higher than its ratings suggest. For example, during the first season of "3-2-1 Contact!", the show was watched at least once each week by 23% of children ages 2 to 11 (Chen, 1984). An unusually high percentage of children who view the show (61%) report doing something as a result of viewing; these activities range from discussing the show to conducting an experiment, visiting a museum, or reading a book (Crane, 1987).

Although public television has a relatively low share of television viewing overall, there is some evidence that programs about science, animals, or nature are a highly valued part of the public television schedule. A survey of over 1,000 adults in California showed that more respondents (59%) would miss these programs "some" or "a lot" than public television programming in any of a dozen other categories (Field Institute, 1985).

The audience for public television is diverse. For example, in October 1985, 54% of the adults viewing television in the previous 7 days reported household income under $25,000, compared to 24% with income over $35,000. Fifteen percent of the viewers were black or Hispanic (CPB, 1986b). In the California survey referred to above, 60% of the respondents disagreed with the assertion that "the programming on my local public television station is designed to appeal primarily to highly educated people" (Field Institute, 1985).

Television Funding

In 1985, the estimated expenditures for advertising on commercial television were nearly $21 billion. By comparison, total income for public broadcasting in 1985 (including radio, which is a small fraction of the total) was about $1.1 billion. During the decade 1975-1985, there was a dramatic decline in the proportion of public-broadcasting income received from the federal government, from 25.3% to 16.3%; state and local government's share also declined. The nongovernment share of the total increased from 31.8% in 1975 to 50.7% in 1985 (U.S. Dept. of Commerce, 1987).

Commercial and Public Television in Perspective

An important aspect of science on television is the very great importance of commercial, nonscience television programs, such as dramatic series and the news, in conveying information about science. Although these information sources are obviously difficult to influence, they cannot be ignored in any discussion of the transmission of science information (and misinformation) to the public. During certain periods, for example, science-related topics dominate television news (e.g., space exploration, the Challenger explosion, Chernobyl)--and television is now the major source of news for most people.

The relatively small amount of viewing time devoted to public television (1 hour in 20) does not diminish the importance of science on public television (which the public feels is very important, according to the Field poll), but does place this information source in some perspective with respect to other sources of science information. The 5% figure also leads rather quickly to considerations of who watches public television, why, and what they get from it--a complex subject beyond the scope of this analysis, except to note that for a substantial segment of the population--those segments of the public "attentive to" or "interested in" science policy--public television science series are the most popular television shows.

American Museums as a Source of Informal Science Education

To understand the capabilities of science museums as a medium of informal science education, one needs to understand the numbers and diversity of these institutions and where they derive their support. According to a 1979 survey, there were at that time 4,400 nonprofit museums in the United States; some knowledgeable observers put the figure at about half that number.* History museums were most common, constituting about half of all museums. Science museums were next most common; the 800 science museums made up 18% of the total (NCES, 1979).

The numbers of science museums are apparently increasing, although it is difficult to establish the trend with any precision, given the problems in defining "museums" and cataloguing them. By 1986, the number of museums listed in the *Official Museum Directory* had grown to about 6,050 (excluding aquaria, arboretums, botanical and aquatic gardens, conservatories, and zoos). Of these, about 1,550 (or 26%) were categorized as science museums; most of these are small. If nature centers as well as park museums and visitor centers are included, the total rises to about 2,100 museums (or 35% of the total) (American Association of Museums, 1986).

By comparison with other museums, science museums require a large amount of resources. The data available for 1979 indicate that 37% of the total operating expenses for all museums went for operation of science museums--a figure more than double the proportion of science museums reported at that time (18%). A total of $368 million was spent to operate the science museums (NCES, 1979).

---

* Not all members of the science museum community put great stock in these figures.

Science museums draw on a diverse array of public and private resources. In 1979, science museums' operating income was obtained as follows: earned income, 42%; federal government, 14%; state government, 6%; local government, 25%; individual contributions, 4%; foundations, 5%; corporations, 2%; and other, 3%. This pattern was very similar to the support for all museums, except that science museums' reliance on state government was lower than for most museums (6% vs. 12%), while reliance on local government was higher (25% vs. 17%) (NCES, 1979).

- To some degree, an infrastructure exists within the science museum field through the existence and growth of associations representing museums. The American Association of Museums (AAM) includes museums of all types as institutional members and also has a large individual membership, most of it consisting of persons working for museums. The annual *Official Museum Directory*, published by AAM, includes a comprehensive list of American museums.

- The Association of Science-Technology Centers (ASTC) includes institutional members only, in five membership categories. As of September 1987, ASTC had 226 members located in the United States and 22 other countries. ASTC provides a wide variety of services; for example, it has established a traveling exhibit service operating on a national scale, specializing in hands-on science exhibits, and it initiated the Science Teacher Education at Museums (STEAM) program funded by the General Electric Foundation, which provides museums with funds for teacher education.

Visitors

On the basis of a 1972 study, science museums appear to be visited in numbers far out of proportion to their representation in the museum population. About 38% of all museum visits in that year were to science museums, whereas only 16% of museums were science museums (National Endowment for the Arts, 1974). The 1979 data show a similar proportion: 150 million of the total of 348 million museum visits in that year were to science museums; that is, about 43% of all visits were to the 18% of museums that were science museums (NCES, 1979). Large museums account for the great majority of all museum visits. In 1972, 45% of the visits were accommodated by the 10% of museums with the largest operating income (National Endowment for the Arts, 1974). (This fact leads to the inference that growth in the number of science museums, per se, is not likely to cause a proportional growth in the number of visitors, since most new museums will be small.)

Visitors to science museums are about half children (through age 17) and half adults. A 1971 survey found that 49% of visitors to science museums were in grades 1-12, while another 14% were in college (McGrath, 1971). Similar figures were found for combined history/science museums; for example, 52% of the visitors were in grades 1-12. Anecdotal reports suggest that these proportions are still about the same, that is, 50-50 for children and adults.

For adults, at least, museum visits are discretionary. Not surprisingly, many adults visit museums infrequently, if at all. For example, one New York City survey found that 62% of respondents indicated that they visit *any* type of museum only once a year, or less often; that is, only 38% visit a museum more often than once a year (*Newsday*, 1985). Data on the "attentive" public for science (representing about 20% of the population) show that in 1981, 52% of them had visited a science or technology museum in the preceding year, whereas among the "interested" public (the next most interested 20% portion of the population), only 28% had done so (Miller, 1983).

There are no good national data that enable us to describe the characteristics of museum goers. The only research we could locate that pertains to this topic was too old and localized (in one city) to be useful for purposes of this macro view.*

*Impact***

Impact is very difficult to measure, and only limited attempts have been made to do so (Diamond, 1987). This is true for all types of museums. Nonetheless, research has shown that, at one extreme, for students who worked in a science museum as "explainers," the effects up to 15 years later were "strong and persistent" (Diamond, 1987).

Opinions of museum directors provide some clues to the possible impacts of their museums on visitors. The 1972 museum survey found that a higher proportion of directors of science museums than of any other type of museum considered "providing educational experiences for the public," and also "providing instruction to the young," to be very important to their mission (National Endowment for the Arts, 1974).

---

* A 20-year-old Smithsonian survey shows that visitors to the National Museum of Natural History and the National Museum of History and Technology were predominantly well educated; 43% were college graduates. Also, 32% of those visitors were in the professional and semiprofessional categories, and another 27% were married to such persons (Wells, 1970).

** A separate pilot activity explored in depth the issues related to the effect of informal science learning experiences on the individual (Sections VI and VII). For reasons that our subsequent discussion of this activity makes clear, there is little available information on this topic with which to construct an aggregate profile.

For museums, as for all informal educational institutions, entertainment and education are closely intertwined. For example, a study of the National Air and Space Museum found that 57% of the visitors said their motive for the visit was primarily the desire for pleasure and entertainment, compared with 36% who came primarily for educational purposes (Yankelovich et al., 1977).

The Quality of Data on Science Museums

The age of much of the data about science museums weakens any attempt to construct a meaningful statistical profile. More recent data may be different in significant ways; for example, the demography of individuals visiting science museums (age, race, sex, education, etc.) may reflect more aggressive outreach efforts in recent years.

The field suffers from definitional problems--for example, what is a "science museum"? For certain purposes, nature centers as well as park museums and visitor centers are considered to be distinct from science museums. Yet, in terms of the numbers of visits and what is learned, the latter may be quite important components of the informal science education scene. On the other hand, AAM includes zoos, aquaria, botanical gardens, and the like as science museums, whereas many other surveys and organizations do not. Clarity about definitions is more important in discussing the field of science museums than one would suppose; however, most writers seem rather casual in their use of the term "science museum."

Efforts are under way to remedy these problems, in part as the result of NSF's efforts.

- ASTC is currently completing work on a survey of the science-technology museum field. The survey (which is funded by the National Science Foundation) will provide information on many topics, including annual attendance, interior square footage, exhibitions, and education programs.

- For the first time, AAM is including a brief questionnaire with its annual update forms for museums; however, when these data may become available is unknown. A federal agency, the Institute for Museum Services, expects to publish soon the results of a survey of federal agencies focusing on their support for museums.

- The Association of Systematics Collections (ASC), which consists of the 80 largest natural history museums in the United States, is collecting annual reports from its member institutions and plans to compile some of the information in them, such as annual attendance figures, in preparation for an NSF-supported conference being organized by the ASC for next October.

Until such data as these are made available, it is difficult to compile an accurate picture of science museums and their place in the larger community of museums. NSF may wish to stay abreast of efforts to update statistical information on museums.

REFERENCES

Overview

CPB (Corporation for Public Broadcasting). (1986a). Informal, out-of-school learning in the United States: Who, what, how, why? *Research Notes*, No. 18 (May). Washington, DC: Author.

CPB (Corporation for Public Broadcasting). (1982). *Public television programming content by category, fiscal year 1982.* Washington, DC: Author.

Evered, D., & O'Connor, M. (Eds.). (1987). *Communicating science to the public (Ciba Foundation Conference).* New York: John Wiley & Sons.

Field Institute. (1985). *The Field Institute/The California Poll, early May 1985 (Data File #8502).* San Francisco: Author.

Hueftle, S. J., Rakow, S. J., & Welch, W. W. (1983). *Images of science: A summary of results from the 1981-82 National Assessment in Science.* Minneapolis, MN: University of Minnesota.

IEA (International Association for the Evaluation of Educational Achievement. (1988). *Science achievement in seventeen countries: A preliminary report.* New York: Pergamon Press.

Juster, F. T. (1986). Preferences for work and leisure. *Economic Outlook USA*, 1st quarter 1986, pp. 15-17.

Laetsch, W. M. (1987). A basis for better public understanding of science. In D. Evered & M. O'Connor (Eds.), *Communicating science to the public (Ciba Foundation Conference).* New York: John Wiley & Sons.

Miller, J. D. (1986). *Some new measures of scientific illiteracy.* Paper presented to the 1986 annual meeting of the AAAS, Philadelphia.

Miller, J. D. (1987a). Scientific literacy in the United States. In D. Evered & M. O'Connor (Eds.), *Communicating science to the public (Ciba Foundation Conference).* New York: John Wiley & Sons.

Miller, J. D. (1987b). *The roots of scientific literacy.* A paper presented to the 1987 St. Louis Sigma Xi Symposium on Science Education.

Murnane, R. K., & Raizen, S. A. (Eds.) (1988). *Improving indicators of the quality of science and mathematics education in grades K-12* (page 8). Washington, DC: National Academy Press.

Overview (Continued)

NCES (National Center for Education Statistics). (May 1984). *Science and mathematics education in American high schools: Results from the High School and Beyond Study.* Washington, DC: Author.

NSB (National Science Board). (1985). *Science indicators: The 1985 report.* Washington, DC: U.S. Government Printing Office.

NSB (National Science Board) Commission on Precollege Education in Mathematics, Science and Technology. (1983). *Educating Americans for the 21st century.* Washington, DC: U.S. Government Printing Office.

Riccobono, J. A. (1986). *Use of electronic information technologies for non-school learning in American households (report of findings from the 1985 Home Information Technology Study [HITS]).* Washington, DC: U.S. Department of Education, Center for Statistics.

Traub, J. (1985). The world according to Nielsen: Who watches television--and why. *Channels* (January/February).

U.S. Department of Commerce. (1980). *Social indicators III: Selected data on social conditions and trends in the United States.* Washington, DC: U.S. Government Printing Office.

U.S. Department of Commerce, Bureau of the Census. (1987). *Statistical abstract of the United States, 1987.* Washington, DC: U.S. Government Printing Office.

U.S. Department of Education. (1986). *The condition of education.* Washington, DC: U.S. Government Printing Office.

Television

Chen, M. (1984). *A review of research on the educational potential of 3-2-1 Contact: A children's TV series on science and technology.* A report prepared for the Children's Television Workshop and the NSF.

CPB (Corporation for Public Broadcasting). (1986a). Informal, out-of-school learning in the United States: Who, what, how, why? *Research Notes*, No. 18 (May). Washington, DC: Author.

CPB (Corporation for Public Broadcasting). (1986b). *Summary audience data for public television and radio.* Washington, DC: Author.

Crane, V. (1987). *An exploratory study of "3-2-1 Contact" viewership.* Chestnut Hill, MA: Research Communications, Ltd.

Television (Continued)

Field Institute. (1985). *The Field Institute/The California Poll, early May 1985 (Data File #8502).* San Francisco: Author.

Gerbner, G., Gross, L., Morgan, M., & Signorelli, N. (1981). Scientists on the TV screen. *Culture and Society* (May/June).

Miller, J. D. (1983). Reaching the attentive and interested publics for science. In S. M. Friedman, S. Dunwoody, & C. L. Rogers (Eds.), *Scientists and journalists: Reporting science as news.* New York: Free Press (Macmillan).

NSB (National Science Board). (1985). *Science indicators: The 1985 report.* Washington, DC: U.S. Government Printing Office.

Traub, J. (1985). The world according to Nielsen: Who watches television--and why. *Channels* (January/February).

Tressel, G. W. (1984). Supporting science: A stand against lowest-common-denominator TV. In K. Hays (Ed.), *TV, science and kids: Teaching our children to question.* Reading, MA: Addison-Wesley.

U.S. Department of Commerce, Bureau of the Census. (1986). *Statistical abstract of the United States, 1987.* Washington, DC: U.S. Government Printing Office.

U.S. Department of Health and Human Services, National Institute of Mental Health. (1982). *Television and behavior: Ten years of scientific progress and implications for the eighties (Summary Report).* Washington, DC: Author.

Washington Post. (1987, November 18). TV ratingzzzz.

Williams, P., Haertel, E. H., Haertel, G. D., & Walberg, H. J. (1982). The impact of leisure-time television on school learning: A research synthesis. *American Educational Research Journal*, 19:1 (Spring).

Museums

American Association of Museums. (1986). *The official museum directory, 1987.* New York: National Register Publishing Co., Inc.

ASTC (Association of Science-Technology Centers). (1980). *Exploring science: A guide to contemporary science and technology museums.* Washington, DC: Author.

Diamond, J. (1987). *Learning through teaching: Long-term impacts of science museums.* Paper presented at the Annual Meeting of the AAAS, Chicago.

Museums (Continued)

Dunbar, N. G., & Borun, M. (1980). *The science museum audience.* Washington, DC: Association of Science-Technology Centers.

Evered, D., & O'Connor, M. (Eds.). (1987). *Communicating science to the public (Ciba Foundation Conference).* New York: John Wiley & Sons.

McGrath, K. M. (1971). *1971 financial and salary survey.* Washington, DC: American Association of Museums.

Miller, J. D. (1983). Reaching the attentive and interested publics for science. In S. M. Friedman, S. Dunwoody, & C. L. Rogers (Eds.), *Scientists and journalists: Reporting science as news.* New York: Free Press (Macmillan).

National Endowment for the Arts. (1974). *Museums USA.* Washington, DC: Author.

NCES (National Center for Education Statistics). (1979). *Museum program survey.* Washington, DC: Author.

Nichols, S., et al. (Eds.). (1984). *Museum education anthology: Perspectives on informal learning* (esp. section on "First Questions"). Washington, DC: Museum Education Roundtable, American Association of Museums.

Wells, C. H. (1970). *Smithsonian visitor: A survey.* Washington, DC: Smithsonian Institution.

Yankelovich, Skelly and White, Inc. (1977). *Visitors' reactions to the National Air and Space Museum and the individual exhibits. Vol. II: Detailed findings.* New York: Author.

III MARKET ASSESSMENT FOR A NEW INVESTMENT AREA: EXAMINING THE POTENTIAL OF VIDEOCASSETTE TECHNOLOGY AS A VEHICLE FOR INFORMAL SCIENCE LEARNING IN THE HOME

By

Teresa Middleton

Andrew A. Zucker

III MARKET ASSESSMENT FOR A NEW INVESTMENT AREA: EXAMINING THE POTENTIAL OF VIDEOCASSETTE TECHNOLOGY AS A VEHICLE FOR INFORMAL SCIENCE LEARNING IN THE HOME

Part of the job of any NSF program officer is to be alert to whole new areas of investment that have potential for improving science education in and out of the schools. There is a dilemma here, however, for NSF program officers tend to be familiar with fields and media already working with NSF (e.g., science museums, children's science television). It takes time and effort to develop knowledge about new fields and make connections with potential grantees.

As an example, the home and school use of videocassette recorders (VCRs) is now so widespread that it may offer a major new "distribution channel" for science education. There are questions, however, about the production capabilities of the field, the market demand, and the distribution mechanisms for science-related videotapes. Answering such questions, however, involves research that requires resources and expertise that are unavailable to NSF program officers. At NSF's request, we undertook this study to illustrate how a short-term, focused market assessment, contracted to an external party who is an expert in this area, could help to explore the potential for new initiatives.

The objectives of this assessment activity were to (1) make a preliminary assessment of the market for science-related videocassettes for informal learning at home by teenagers and adults, and (2) provide guidelines for conducting a full market study. "Science-related" was broadly defined to include science, mathematics, technology, and nature. The definition of "science" may be even broader for young children, which is one reason why we excluded them from the analysis.

Highlights

Our study, carried out by a marketing expert at SRI, was a scaled-down version of a full market study. The results are therefore more tentative and less fully supported than would be the case in a full market assessment (which could provide valuable assistance to NSF's planning if this investment area appears attractive at some future date). We review the main themes in our findings below.

- The VCR market for home, school, and business is growing very rapidly: from 1985 to 1990 the market is likely to increase sevenfold.

- Currently, health and science topics make up nearly a quarter of existing videocassette titles; however, the volume of sales and rentals is far smaller.

- The market for "home learning" VCR products is embryonic. A few producers are testing the waters; others are waiting to see what will happen.

- A viable market for science education videocassettes targeted to the home (adult and teenage audiences) seems unlikely to emerge in the next 5 years.

- Various factors may alter the future size and shape of the home science learning market, among them, changes in the technology (interactive video), increased school use of VCRs, and decreasing perception that VCRs are "special."

- This preliminary assessment seems to confirm opinions based on other evidence that supporting a home-based videocassette initiative at this time does not appear to be an especially valuable undertaking for NSF if the Foundation's goal is to achieve widespread impact on home science learning (however, NSF may have a legitimate exploratory research and development role in this area).

Methods for Pilot Market Assessment

SRI International has performed market assessments for many clients, and we used established market research procedures for conducting this one. When conducting a full-scale market assessment study, we first define a set of companies and individuals from whom information about the market may be obtained. This is not a random set, but rather we purposefully select those companies and individuals who may be leaders in the industry, or who for any other reason have knowledge of industry dynamics and trends.

The list of companies and individuals serving as data sources is entered into a computerized data base management system and becomes the basis for a tracking sheet (listing companies and individuals to be interviewed, dates of interview, expected results, and actual results). Mailing addresses for letters of introduction and thanks sent to participants are also produced by computer.

During the pilot study we generated a list of some 60 companies and individuals (Exhibit III-1) that could be used in a more complete assessment of the market for home videocassettes for informal science and mathematics learning. We interviewed a subset of these 60. We would recommend approximately 50 interviews be conducted for a full-scale assessment of the market.

For the pilot assessment, we collected preliminary data from the variety of different sources we would tap in a full-scale market assessment. These include:

- Interviews with video developers and publishers, industry watchers, educational institutions, distributors, and customers.

- Publicly available information from SRI's libraries and data bases, and additional information that already resides at SRI in the form of nonproprietary research data.

The purpose of such data collection is not just to assess the existence of a market or to establish its size. It is equally important that we understand the environment of that market and the dynamics of the industry serving it. Thus, in the pilot study, informal interviews were undertaken to help us understand the industry and its products, market size and market trends, competitive dynamics, and technology trends. These interviews were also to confirm information we had gathered from other sources. This process provided us with a set of questions that form the basis of an interview guide that can be used in a larger market assessment (see Exhibit III-2). Such a full-scale assessment might, if warranted, include a broader survey of consumers or producers to test preliminary findings.

Readers should understand that a market assessment, by its nature, necessarily combines facts and other data with judgments about such information.

Pilot Assessment Findings

For the preliminary study we prepared a report outline that would be appropriate for a full market assessment (see Exhibit III-3). The outline has been used for the report on the pilot study also. It focuses not only on the size of the home market for science-related videotapes, but also on market and product environments, competitive dynamics, and technology trends. All affect the future of the industry. The remainder of this section details our pilot findings under each major heading of the assessment report outline: (1) summary: yesterday, today, and tomorrow; (2) market environment; (3) product environment; (4) competitive dynamics; (5) technology directions; and (6) conclusions: market attractiveness.

Summary: Yesterday, Today, and Tomorrow

Videocassette recording, introduced in its present form by Sony in 1975, found ready acceptance in the United States, Asia, and Europe. Today approximately 52% of all American households have at least one VCR; projections are that 85% will have a VCR in 1995. VCRs are already taking viewing share from network and cable, and by 1995 it is estimated they will account for 25% of all viewing time.

Sixty-one percent of VCR-owning households contain more than three people; and in 36% of these households there are children between the ages of 13 and 19. Families are using their VCRs with increasing frequency. On an average, a family of four now runs a VCR for 97 minutes per day.

VCRs are also finding their way in increasing numbers into the schools, where teachers (familiar with using filmstrip and video) find them technically simple to use and a familiar medium to work with for group learning.

VCR sales amounted to $3 billion in 1985 and are expected to increase to $20 billion in 1990. Video software sales and rentals to home consumers, businesses, and educational institutions have kept pace with the rapid installation of hardware. Assuming a conservative $50 was spent in a year on tape rentals and purchases for each machine, the video software market should generate over $700 million this year.

General interest/education and health/science videocassettes dominate the categories of *titles* being distributed. In terms of revenue, movies are dominant.

Market Environment

Industry Structure--The videocassette industry is relatively young; however, most of the key producers are well known to their audiences. They have their roots in film (particularly entertainment), television, educational audiovisual materials (particularly filmstrip), or print publishing.

The market contains three major segments: home, business, and education. Initially, companies focused on discrete market segments and were identifiable by these segments. These providers included film companies (e.g., RCA and Warner), textbook publishers (e.g., Addison-Wesley and Hayden), and educational media companies (Deltak and CRM/McGraw-Hill).

Barriers between market segments have started to blur in the past 5 years, and there has been an explosion of participants in the industry. Universities and religious institutions should now be counted as increasingly strong producers. There also exist a number of small companies, most of them concentrating on how-to videos.

The market for videocassettes to provide informal learning in the home, however, is at the embryonic stage, with few producers.

Market Characteristics: Informal Home Learning--There are similarities between the market for home educational videocassettes and the market for home educational computer software. Educational software companies originally directed their sales to the school market, but there is now considerable overlap in that industry between education and consumer software markets. Most successful companies address both.

Similarly, the dramatic growth of the use of educational videocassettes in schools during the past 5 years is likely to have some influence on their introduction into the home, if only as supplementary materials.

Distribution Channels--Distribution channels have been established according to content of the videotapes as well as to market segments. The major film companies have set up separate home video divisions for "repurposing" (adapting existing film material for video release) and distributing their films and entertainment tapes. Their consumer outlets are franchised video stores; mass merchandisers, including liquor stores and bookstores; and libraries.

Education material is distributed through traditional education film and other audiovisual distributors and dealers. Some instructional and how-to tapes are also starting to trickle through consumer outlets.

Videocassette distribution is unique in that it supports both rentals and sales. The rental strategy, to capture the movie consumer, has worked very well. However, companies are increasingly looking for ways to make renting a tape easier or purchasing more attractive.

Product Environment

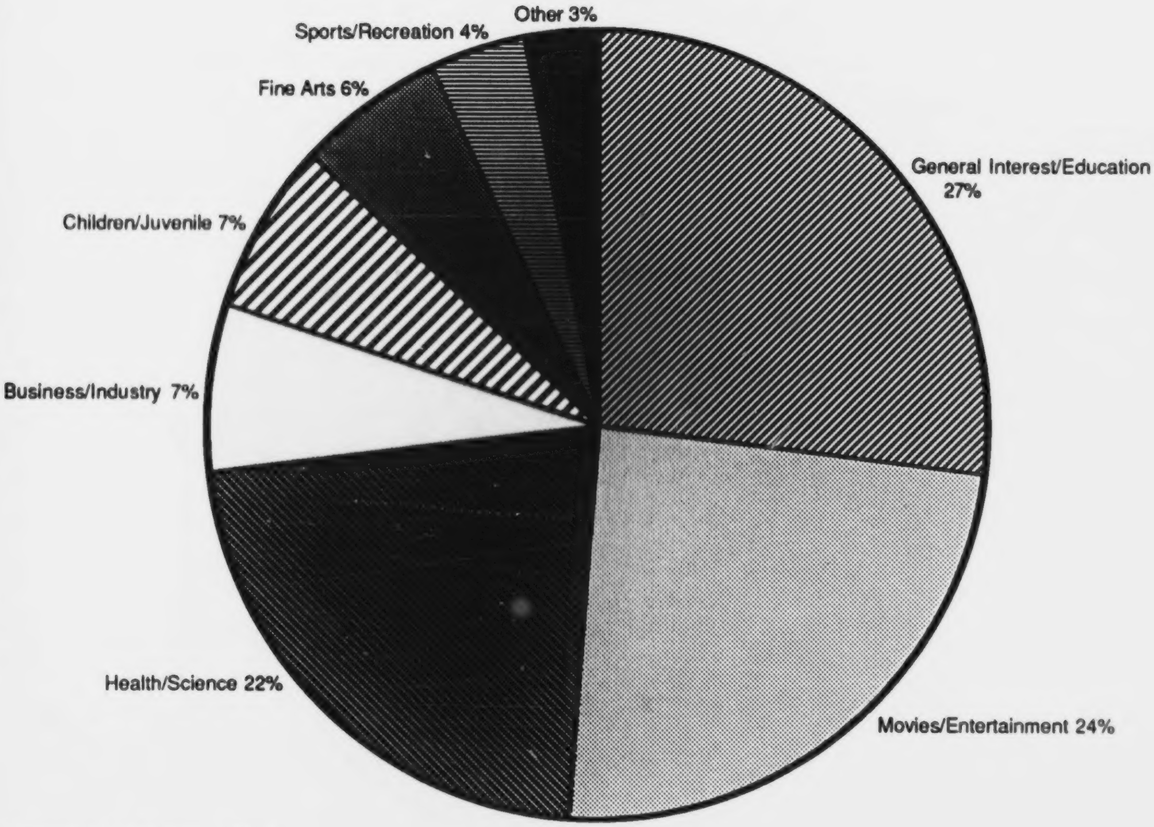
Product Categories and Characteristics--Figure III-1 shows a breakdown of the content of the more than 54,000 existing videocassettes. Surprisingly, movies and other entertainment account for less than 25%; general interest/education and health/science make up 49% of all videocassettes.

If general interest/education and health/science continue their dominance of this medium (in numbers of titles), it can be expected that most sales will be to instructional institutions: elementary and secondary schools, postsecondary colleges and universities, nursing schools and hospitals, and so on. However a small percentage of them will be earmarked for home sales. More than 5,000 of these existing videotapes are science related. Not all are available for distribution, and we estimate that fewer than 5% would lend themselves to repurposing for a home audience. The list of subjects under which these tapes fall (see Table III-1) illustrates how "science related" might be defined.

Sources of Differentiation--The consumer videocassette market segment in terms of sales (compared with rentals) is very small, according to a spokesman for A. C. Nielsen. There is an increase in households willing to purchase, but the number has not grown to the extent that the industry expected with recent price cutting among competitors. This pattern indicates that the market is relatively inelastic--those who want to buy will buy regardless of price.

The few science-related products currently available for home users have little price differentiation. They are priced below $50, which appears to be the "ceiling" price for videotapes for the home purchaser, and may be discounted to around $30.

Traditional video is star and title driven, and preliminary successful products in this small niche market may follow that trend. Successful products are led by those coming from National Geographic, most of which had their beginnings in quality television programs (e.g., "The Incredible Human Machine"). Other products include "Planets of the Sun," starring Leonard Nimoy, and Spinnaker Software Company's "Beat the SAT--Math and Verbal." All three products mentioned have shown up on lists of top special-interest videocassette sales in the business and education categories, and it can be assumed that they are reaching home consumers as well as schools.



(percentage of all current titles)

FIGURE III-1 CONTENT OF EXISTING VIDEOCASSETTES

Table III-1

VIDEOCASSETTES AVAILABLE IN SCIENCE SUBJECT CATEGORIES

Aeronautics
Anatomy and physiology
Animals
Antarctic/Arctic regions
Biology
Birds
Blood
Bones
Chemistry
Childbirth
Circulatory system
Computers
Dinosaurs
Earthquakes
Electricity
Electronics
Embryology
Energy
Evolution
Genetics
Geology
Immunology
Mathematics
Metabolism
Metallurgy
Meteorology
Metric System
Microbiology
Muscles
Natural resources
Neurology
Oceanography
Physics
Science
Scientists
Seasons
Senses
Solar energy
Space exploration
Technology
Volcanoes
Waste products
Wildlife
Zoos

New Product Development--Video is costly to produce. A practical budget for a made-for-video feature (as opposed to repurposing old material) would be in the range of $1 million on 35-mm film, and $500,000 on 16-mm. (Developers are reluctant to shoot on tape because doing so precludes foreign sales.)

Future Product Demand--This topic needs deeper investigation. Preliminary analysis indicates that the market for science and mathematics education tapes for home viewing may be 5 years away. By then, video-viewing habits will be changing, distribution channels will be in place, and the use of video in schools will have influenced home use, if that is to take place.

Public and school libraries that provide free access to videotapes, and rental outlets that supply tapes at affordable prices, will continue to take the steam out of the tape purchase market. People are expected to continue to use home VCRs primarily as entertainment vehicles; however, other motives could influence the entry of science-related tapes into the home during the next 5 years. For example, home-based VCRs could provide support to schooling, and students might be encouraged to view tapes for credit or credentialing. Many tapes that teach science concepts (e.g., space science) might also come under the broad heading of "entertainment" and find their way into the home market.

Competitive Dynamics

Competitive Positioning--At present there is little competition for the market for videocassettes to provide informal learning in the home. However, there are companies (such as National Geographic) who are already testing the market by reworking existing footage (generally television productions) and marketing the resulting product as videotape.

Other companies are waiting in the wings to see whether the market is a viable one. Educational filmstrip companies are looking for products to replace filmstrip, which is coming to the end of its product life, and videotape is a natural for them. Educational and consumer software companies and textbook publishers will also be watching to see whether this market presents a valid diversification strategy.

Driving Forces--The principal driving force in the industry has been the rapid VCR penetration in the home and school markets. However, there is concern that VCRs will become increasingly like TV--"not special any more," as the chairman of a leading university's department of theater arts says. VCR use declines over time, and people depend less on prerecorded tapes for entertainment. This trend needs further investigation; it may not be applicable to education.

Technology Directions

Technology Changes--Sony has accepted VHS as the standard for VCRs, and will discontinue production and sales of its Beta machines, although it will support those already sold. Thus, a technology that was already easy for the consumer to use has become even simpler, and there will be only one standard of tape to select from.

Videocassettes will increasingly become interactive, opening the way for gaming and simulations and thereby supporting delivery of education that is likely to be more saleable to the teenager.

Equipment Suppliers--The list provided in Exhibit III-1 includes the key videocassette developers and distributors in the country.

Conclusion: Market Attractiveness

No conclusion can be reached about the future attractiveness of the market without conducting a full market study; however, it appears unlikely that any sizable market will emerge during the next 5 years.

The Utility of Market Assessments

Market assessments such as this describe the way NSF-supported materials, films, or other products reach intended (or other) audiences. As such, it helps NSF understand whether a particular kind of investment is likely to have a widespread impact on the science learning opportunities for young people. The findings of this assessment--although preliminary--suggest caution on NSF's part, if its goal is to have science-related videocassettes widely used in the home in the near future.

However, the results of this kind of market assessment can be interpreted differently, depending on how the Directorate chooses to position itself in relationship to the cycle of technology development and diffusion. If NSF wishes to advance the frontiers of knowledge and development in this area of learning technology, then the market assessment might be interpreted as evidence of a present opportunity. Under this scenario, even though the technology itself is already developed and proven, prototypes of videocassette programming targeted to teenage years (or other age levels) do not yet exist in great numbers. Similarly, the full possibilities of interactive video have not yet been examined. The argument could be made that NSF has the chance to set the standard or explore the possibilities for this medium. Small-scale R&D investments would probably make sense under this set of assumptions; the success of such investments would not rest on widespread adoption or use.

This kind of exploratory development work can pay off handsomely, as attested to by the development of archiving methods using CD-ROM technology. A Carnegie grant to develop a marketable archive of NSF-supported K-6 elementary science programs and materials (produced over a quarter of a century) has in the last year led to a feasible product for which there is considerable demand. The principal investigator observed:

> **On the CD-ROM project, we would have been dead in the water if we made our decision based on the installed CD-ROM machines or knowledge of them. When a field is first opening, there is risk.... In our case with the CD-ROM, it appears that the gamble has worked. The disc is out, sales are growing in a logarithmic fashion; we had many publishers fighting to produce the product; and now Computerland, which will sell Amdek players, is choosing to bundle the Science Helper K-8 disc with its machine. And the country gets to recover some further value from its initial investment in science curriculum development.**

If, on the other hand, NSF decides it is in the business of reaching mass audiences in the short term--an emphasis of its current investments in informal science education--then investments designed to "seed" the home videocassette market may be premature.

The important point is that the market assessment, by itself, does not--and cannot--provide definitive direction for investment decisions. Market assessments are one tool that can help NSF's staff decide whether an initiative aimed at a particular "market niche" is or is not likely to be viable. The decisionmaking process would also consider a variety of other information, such as potential costs and benefits, political viability, staffing requirements, and so on. In combination with other assessment findings (e.g., concerning patterns of school videocassette use, informal learning from video sources, and teenage leisure-time science activities) and clear goals for the Foundation's role in this domain, market assessment can help NSF planners focus their resources productively.

REFERENCES

Billboard, International Newsweekly of Music and Home Entertainment (1987 issues).

Electronic Industries Association, quoted in *San Francisco Chronicle* (1988, January 12).

Hayes, J. (Ed.). (1986, May). *Video purchasing patterns in schools.* Denver, CO: Quality Education Data.

Leisure Time Electronics. (1986, January).

Link Resources Corp. (1986, April). *Preliminary findings, 1986 consumer survey, VCR owner segment*. New York: Author.

National Video Clearinghouse, Inc. (1987). *The video source book* (9th ed.). Syosset, NY: Author.

Wilkofsky Gruen Associates. (1986, March). Quoted in *Market Monitor*, published by the International Communications Industries Association.

Exhibit III-1

LIST OF COMPANIES TO PROVIDE INFORMATION FOR ASSESSMENT OF THE HOME VIDEOCASSETTE MARKET FOR INFORMAL SCIENCE AND MATHEMATICS LEARNING

Addison-Wesley Publishing Co.
Route 128
Reading, MA 01867
617-944-3700

Aero/Space Visuals Society
151 Farmington Avenue
Hartford, CT 06156
203-273-0123

American Bible Society
1865 Broadway
New York, NY 10023
212-581-7400

American Video Tape
1116 Edgewater Avenue
Ridgefield, NJ 07657
201-941-4404

Anderson/Hickey Productions
4207 Leewood Road
Stow, OH 44224
216-686-1628

Churchill Films
662 North Robertson Boulevard
Los Angeles, CA 90069
213-657-5110

Coronet/MTI Film & Video
108 Wilmot Road
Deerfield, IL 60015-9990
312-940-1260

CRM/McGraw-Hill Films
674 Via de la Valle, P.O. Box 641
Del Mar, CA 92014
619-453-5000

Crossroads Video
15 Buckminster Lane
Manhasset, NY 11030
516-365-3715

Crown Video
225 Park Avenue S.
New York, NY 10003

Deltak Inc.
1751 Diehl Road
Napierville, IL 60566
312-369-3000

Downstream Science Productions
4377 Carter Trail
Boulder, CO 80301

Educational Materials & Equip. Co.
P.O. Box 17
Pelham, NY 10803
914-576-1121

Environmental Video Inc.
1116 8th Street, "B", P.O. Box 577
Manhattan Beach, CA 90266
213-546-4581

Films Inc.
1075 East Meadow Drive
Palo Alto, CA 94303
415-494-3701

Gessler Educational Software
900 Broadway
New York, NY 10003
212-673-3113

Global Village
454 Broome Street
New York, NY 10013
212-966-7526

Hayden Book Company
10 Mulholland Drive
Hasbrouck Heights, NJ 07604
201-393-6353

Exhibit III-1 (Continued)

Heinemann Educational Books, Inc.
70 Court Street
Portsmouth, NH 03801
603-431-7894

Houghton Mifflin Company
2 Park Avenue
Boston, MA 02108
617-878-2600

International Video Bible Lessons
P.O. Box 2255
West Monroe, LA 71291
318-396-6265

Josten's Educational Services
810 Elm Street, P.O. Box 796
Owatonna, MN 55060
507-455-6100

King Features Entertainment
235 East 45th Street
New York, NY 10017
212-682-5600

Knopf Video Books
201 East 50th Street
New York, NY 10022
212-572-2103

Learning Corporation of America
108 Wilmot Road
Deerfield, IL 60015-9990
312-940-1260

Life Video Gospel Association
P.O. Box 395, 1435 Central Avenue
College Place, WA 99324
509-522-0784

Lorimar Home Video
17942 Cowan Avenue
Irvine, CA 92714
714-474-0355

Massachusetts Institute of Technology
77 Massachusetts Ave, Dept 99, Rm 9-234
Cambridge, MA 02139
617-253-7444

Mastervision
969 Park Avenue
New York, NY 10028
212-879-0448

MCA Home Video
70 Universal City Plaza
Universal City, CA 91608
818-777-4300

MGM/UA Home Video
1350 Avenue of the Americas
New York, NY 10019
212-408-0600

Mindscape, Inc.
344 Dundee Road
Northbrook, IL 60062
312-480-7667

Mitchell Publishing Inc.
915 River Street
Santa Cruz, CA 95060
408-425-3851

MTI Teleprograms Inc.
108 Wilmot Road
Deerfield, IL 60015-9990
312-940-1260

National Geographic Society
17th & M Streets, N.W.
Washington, DC 20036
202-857-7378

PBS Video
1320 Braddock Place
Alexandria, VA 22314-1698
703-739-5380

Perspective Films & Video
65 East South Water Street
Chicago, IL 60601
312-977-4000

Quadras Media Ministry Inc.
128 Kishwaukee Street
Rockford, IL 61104
815-987-3970

Exhibit III-1 (Concluded)

Random House Home Video
201 East 50th Street
New York, NY 10022
212-572-2778

RCA/Columbia Pictures Home Video
3500 West Olive Avenue
Burbank, CA 91505
818-953-7900

Scholastic Lorimar
17942 Cowan Avenue
Irvine, CA 92714
714-474-0355

Signs of Life Films
524 Cascade Drive
Mill Valley, CA 94941
415-383-3680

Simon & Schuster Video
108 Wilmot Road
Deerfield, IL 60015-9990
312-940-1260

Sterling Educational Films Inc.
241 East 34th Street
New York, NY 10016
212-683-6300

Sunburst Communications
39 Washington Avenue
Pleasantville, NY 10570
914-769-5030

Telmar Communications Inc.
902 Broadway, 3rd Floor
New York, NY 10010
212-460-9000

Time-Life Video
1271 Avenue of the Americas
New York, NY 10020
212-484-5940

Vestron Inc.
P.O. Box 4000
Stamford, CT 06907
(203) 323-8900

Video Bible Library Inc.
Box 17515
Portland, OR 97213
206-892-7707

Video Instructional Programs, Inc.
521 Fifth Avenue
New York, NY 10175
212-953-2480

Video Knowledge Inc.
29 Brambel Lane
Melville, NY 11747
516-367-4250

Videolearning Systems Inc.
354 West Lancaster Ave., Suite 105
Haverford, PA 19041
215-896-6600

Voyager Press
2139 Manning Avenue
Los Angeles, CA 90025
213-475-3524

Walt Disney Educational Media Co.
500 South Buena Vista Street
Burbank, CA 91521
800-423-2555

Walt Disney Home Video
500 South Buena Vista Street
Burbank, CA 91521
818-840-1111

Warner Home Video Inc.
4000 Warner Boulevard
Burbank, CA 91522
818-954-6000

Wilderness Video
P.O. Box 2175
Redondo Beach, CA 90278
213-542-5813

Exhibit III-2

SUGGESTED QUESTIONS TO BE ASKED OF PARTICIPANTS IN MARKET ASSESSMENT

1. From your perspective, is the home market for educational videocassettes an attractive one? Within that market, what about the science/ mathematics niche?

2. What do you think the size of that market niche is? What will it be five years from now?

3. Who are the key players? What are the bases for competition in the market, now and in the future?

4. Who, in your judgment, is making the best video for this market? What makes it the best?

5. What do you see as emerging educational issues and technology trends?

6. Can you tell me something about design and production? Are many tapes being repurposed from existing material or do developers start from scratch? About how much would an average tape cost to develop?

7. What do you think is the most profitable way to distribute to this niche of the home market?

8. What role do you see VCRs playing in the next 5 years?

9.* Do you believe there is any special role for local, state, or federal government agencies in stimulating the market for home-based science/math educational videocassettes? If so, what would that role be?

---

* Because of the nature of the client--a government agency--we believe it may be useful to ask this question, although it is not part of our normal procedure for conducting market assessments.

Exhibit III-3

OUTLINE FOR MARKET ASSESSMENT REPORT

I *Summary: Yesterday, Today, and Tomorrow*

II *Market Environment*

- A. Industry Structure
- B. Market Characteristics: Informal Home Learning
- C. Distribution Channels

III *Product Environment*

- A. Product Categories and Characteristics
- B. Sources of Differentiation
- C. New Product Development
- D. Future Product Demand

IV *Competitive Dynamics*

- A. Competitive Positioning
- B. Driving Forces

V *Technology Directions*

- A. Technology Changes
- B. Equipment Suppliers

VI *Conclusion: Market Attractiveness*

IV A CROSS-PROGRAM PRINCIPAL INVESTIGATORS' MEETING: EXAMINING INVESTMENTS THAT ESTABLISH LINKAGES BETWEEN INFORMAL EDUCATION INSTITUTIONS AND THE SCHOOLS

By

Michael S. Knapp

IV A CROSS-PROGRAM PRINCIPAL INVESTIGATORS' MEETING: EXAMINING INVESTMENTS THAT ESTABLISH LINKAGES BETWEEN INFORMAL EDUCATION INSTITUTIONS AND THE SCHOOLS

Alongside its investments in children's television, exhibit development for science museums, and other forms of informal science education, the National Science Foundation has supported a number of projects that create or encourage linkages between informal science education institutions and the schools. These projects begin to address an important opportunity for strengthening school science programs. By examining the experience of these projects, NSF can gain clues to productive avenues for future investment.

To pool experiences of people engaged in this kind of activity, SRI convened a group of SEE staff and selected principal investigators whose projects have fostered different forms of linkage between formal and informal educational institutions. The meeting was also meant as an illustration of cross-program activities that perform an informal assessment and planning function.

An important caveat is in order: the meeting explored *NSF-funded* linkages primarily rather than all possible linkages between informal institutions and the schools. This was a natural artifact of the people who were in the room; the composition of the meeting thus led the discussion to address the issues somewhat more narrowly than might have been the case if a different cross-section of informal institutions had been represented. If NSF wished to pursue this area of investment further, it would need to supplement the results of this meeting with other sorts of information (e.g., from literature reviews, conversations with experts).

Highlights

The meeting identified a range of possible linkages between informal education institutions and the schools. Regarding significant barriers, promising "entry points," and caveats in the effort to establish linkages, the following themes ran through the discussion:

- Although a wide range of linkages have been created, the full potential for partnership between informal education institutions and schools has yet to be exploited. There are strong forces that make it hard to form such partnerships. The two represent different world views on education and science; curricular or instructional policies often constrain informal modes of learning in the schools. These and other factors (e.g., logistical problems, limited capacity of the institution's facility) must be dealt with if viable linkages are to become established.

- The attractiveness of different ways to establish linkages depends on how the informal education institution conceives its role vis-a-vis the schools--for example, as a safe haven for professional development or as a proactive agent for change in school science programs.

- Curriculum development, long-term working roles for teachers and students in the informal institution, and various forms of teacher support offer good possibilities for linkage development. In addition, the informal institutions' position between schools and the scientific world opens up the possibility of significant brokering roles.

- Linkages are more easily formed with schools that do not have large numbers of disadvantaged students; special effort is needed to reach that segment of the student population.

- Relationships with the schools are limited by (1) the nature of an informal institution's "collection" of materials (which may not be transportable) and (2) by the expertise of its staff (who may lack depth in teacher education, for example).

Our analysis of the meeting points to the following conclusions regarding NSF's future investment in this area:

- Without targeting specific types of programmatic solutions, NSF might consider alternative investment targets related to linkage formation--for example, network development among informal institutions themselves, informal education institutions in intermediary roles.

- Current SEE programs can--in principle--support virtually any form of linkage project. But NSF's present ways of soliciting this kind of work do not signal that linkage formation is a central goal and hence may not yield the right mix or quality of projects. SEE should clarify the relative importance of this area in its strategic planning.

- Apart from any changes in its outreach or program solicitation strategy, NSF could promote more aggressively the formation of linkages and even awareness of the concept itself.

Design of the Meeting

The 1-day meeting took place at SRI's Washington, D.C., office in February 1988. Before presenting the findings, we review the design of the meeting; a brief discussion of the methodological lessons learned from the meeting appears in Volume 1 (Section IX) of this report.

Our goal in selecting participants was to represent the range of projects funded by NSF in the last 5 years that have contributed to this investment area and to include all the SEE programs currently or potentially supporting such projects. We also tried to represent a variety of settings in which investments in linkages might take place; we recognized, however, that most linkage projects to date have taken place within large urban areas (where major informal education institutions are situated). In the interest of maximizing the number of sites represented while keeping the numbers attending the meeting to a manageable size, we decided not to include representatives of the schools.

Other criteria figured into the choice of participants. For example, we looked for individuals with recognized standing in the informal science education field, who were articulate and thoughtful, and whose perspectives were likely to differ from one another.

The resulting group of participants included individuals representing:

- Informal educational institutions (and corresponding school systems) in urban, suburban, and rural settings, located in the West, South, and Northeast.
- Projects that had undertaken the following kinds of linkage: teacher training arrangements of various kinds, teacher support networks, institutional linkages, materials development for use by the schools, special programs for students on the premises of the informal institution, and community outreach.
- Different types of institutions: science museums, museum-based networks, zoos, arboretums, and children's museums (not solely focused on science activities).
- Six SEE programs: Instructional Materials Development, Teacher Enhancement, Teacher Preparation, Science and Mathematics Education Networks, Informal Science Education, and Private Sector Partnerships to Improve K-12 Science and Mathematics Education.

In total, eight individuals from informal science education institutions and five SEE staff attended the meeting. A list of meeting participants appears in Table IV-1. We did not compensate participants for their time (however, we did reimburse them for travel expenses).

The meeting concentrated on the following questions:

- In what ways can informal science education institutions provide a resource to school-based science education (and vice versa)?
- To what extent can informal science learning modes be transported into the school setting?

Table IV-1

PARTICIPANTS IN A MEETING TO EXAMINE INVESTMENTS LINKING INFORMAL SCIENCE EDUCATION INSTITUTIONS AND THE SCHOOLS

Informal Education Institutions	
Annette Berkovits	New York Zoological Society
Peggy Cole	New York Hall of Science
Robert Cook	Cornell Arboretum
Whitman Cross	Red Mountain Museum
Kay Davis	Fernbank, Inc.
Frank Gardner	Museum Institute for Teachers of Science
Wayne Ransome	Franklin Institute
Bernard Zubrovski	Boston Children's Museum
National Science Foundation	*(SEE Programs)*
Alan McClelland	Private Sector Partnerships to Improve K-12 Science and Mathematics Education
Alice Moses	Instructional Materials Development
Ethel Schultz	Teacher Enhancement
Susan Snyder	Science and Mathematics Education Networks; Teacher Preparation
Michael Templeton	Informal Science Education

- What dangers are there in creating linkages between the two kinds of institutions? How can these dangers be minimized or avoided?

- What forms of linkage present NSF with its greatest opportunities for improving young people's science learning?

- Are there implications for the way SEE programs coordinate investment efforts and strategies?

The meeting generated a range of ideas about the possibilities for fostering linkages between informal science education institutions and the schools. We review the findings below as follows. First, we describe the range of existing linkages. Following that, we examine barriers to linkage that must be overcome and promising entry points for establishing stronger relationships between formal and informal education institutions. Next, we review several caveats regarding the formation of linkages. In conclusion, we interpret the implications of these findings for NSF's future investments.

Range of Existing Linkages

The project sites represented among the meeting participants exhibit a diverse array of connections between informal education institutions and the schools, far richer than one might suppose from knowing the NSF-funded projects' goals. These linkages take a number of forms; we describe the range of linkage types below, noting examples from the projects conducted by meeting participants.

The first three types of linkage emphasize activities that take place on the premises of the informal education institution and that take advantage of the institution's physical facility (zoo collection, hands-on learning center, etc.).

(1) *Making the informal institution's resources available to organized groups of school children.* This type of linkage happens naturally at most of the institutions as school groups make use of the museum, zoo, arboretum, etc., facilities during school hours. Institutions vary in the degree to which they structure the student groups' time while on-site and also in the extent to which the students' teachers are involved. One science museum, for example, organizes workshops for student classes; a zoo has developed a "jungle laboratory" for visiting groups of students. One science museum even took on the responsibility of providing the equivalent of a full course of instruction for high school students during one school term.

(2) *Encouraging school personnel and students to assume working roles within the informal institution.* Several of the institutions represented in the meeting hire teachers on a part-time basis as "museum associates" who develop materials and workshops, assist with exhibit design, or conduct workshops for visiting groups of students and school personnel. Under

72

graduate high school students have also been taken on in several instances as docents or "explainers."

(3) *Providing training or support for teachers at the informal institution.* In various ways, the institutions design and conduct training and support activities in their facilities that are intended to strengthen teachers' science backgrounds, motivation for further science learning, and instructional capacities. Training takes various forms, each intended to use the institutional facility as a source of scientific or instructional ideas; these activities are typically organized as Saturday morning workshops, after-school sessions, or summer institutes; other forms have been tried, such as periodic "camp-ins" in the science museum.

Two additional types of linkage aim more directly at the curriculum, instruction, and educational programs of the schools. Here, the informal education institution concentrates on the activities that take place on school premises and are designed to work within the current constraints on school science programs.

(4) *Sending staff from the informal institution to work with teachers or classes on school premises.* In several instances, museums or other institutions organize "road shows" in which museum personnel travel to the schools with materials, demonstrations, or activities for students to do. These are sometimes done as demonstration lessons taught by staff from the informal institution or as an adjunct to science classes taught by school teachers themselves. One participant in our meeting regularly spends time in schools trying out ideas for new exhibits or science activities with groups of students.

(5) *Developing materials, kits, etc., for use in school science and mathematics programs.* The materials or curriculum can be the focus of linkage activities as well. One museum has developed elementary science kits and distributed them in sufficient numbers so that all fourth graders in a large urban school district now have them; further kit development aimed at other grades is now in progress. In other cases, a zoo and an arboretum are developing curriculum materials that will be published commercially. Although clearly related to the institution's collection or facility, materials of this sort deal more generically with informal science learning in a particular topical area such as plant ecology or wildlife conservation biology.

A sixth type of linkage occurs at an institutional level. Here, informal education institutions develop relationships, typically long term, that involve the sharing of institutional resources and decisionmaking.

(6) *Establishing formal budgetary and policy linkages between informal institutions and the schools.* In one instance, the informal institution (a science museum) is considered part of a large urban school system and

> receives all of its funding from the school system. In other cases, institutions have developed close, top-level relationships with administrators and policymakers in the schools and have attempted to coordinate activities through this mechanism.

Although their NSF-supported projects had typically been designed with a particular focus, such as the enhancement of teachers' motivation and skills or the improvement of materials, most of the meeting participants had developed more than one kind of linkage. For example, one science museum and a large urban school district had developed multiple linkages that included visits to the museum by class groups of students; a massive kit development and distribution program; teacher training aimed at various groups, but with an emphasis on the teaching staff in nine targeted high schools; teacher overnight camp-ins at the museum; a museum-on-the-mall program, designed to attract people in nonschool settings outside of the museum; and close relationships with school policymakers in efforts to devise and mount a comprehensive overhaul of science programs throughout the city's schools. This case represents the more active end of a continuum; more typically, institutions concentrate their energy on a few types of linkage, while remaining open to other possibilities that might arise.

Barriers and Entry Points

Certain constraints and barriers must be overcome if informal education institutions and the schools are to find common ground. At present, these factors mean that partnerships in efforts to improve science education are not easily formed. Participants in the meeting identified both the most significant barriers and a series of promising entry points for establishing or extending the linkage with the schools.

Barriers That Must Be Overcome

Several critical barriers to linkage stand in the way of durable and productive relationships between schools and informal education institutions. Perhaps most important, "two cultures" need to be bridged--that is, school people need to appreciate and value informal science learning as a legitimate mode of education, and at the same time, informal-institution people need to appraise more accurately the goals of, and constraints on, the formal educational system.

The two-cultures problem manifests itself in several ways. First, school people have relatively poor backgrounds in science by comparison with most staff in informal science institutions. Not only are most teachers poorly trained, as numerous national analyses have pointed out, but the conditions of teaching tend to isolate school personnel from each other and from ongoing developments within the scientific world. As a result, many school personnel don't have a base of knowledge to support their own further learning from informal science institutions. Second, teachers tend

74

to be less familiar and comfortable with the idea that "playing" informally with scientific materials constitutes a legitimate mode of learning. Such approaches to learning are not generally encouraged by school curricula, nor have most teachers had much experience with this form of learning (or recognized their own informal learning as such). The director of one museum-based teacher training program observed that:

> The teachers [who attended workshops at the museum] came to us as "experts," so they couldn't play. They were the ones that were designated by the schools [as science specialists], and so they were initially defensive.

Third, because it emphasizes "play," there is a tendency for school people not to take informal science education seriously. Visits to a science museum or other institution are too easily perceived as "time off" from the serious business of school, therefore as having little or no important relationship to science education.

The two-cultures problem can exist within the informal institution itself and needs to be addressed there through meaningful incentives for institution-based scientists to become involved. One participant observed after the meeting:

> More often than not, schools are interested in providing students or teachers to act as interns in the institution. This is *not* an example of a meaningful incentive. The training and supervision required are often not worth the investment of time from the scientists' point of view. A more relevant incentive might be the provision of funds either for scientific research, or reimbursement for the scientist's time when he or she is involved with a science education program.

In one way or another, all the participants in the meeting had attempted to build bridges between the two cultures. The bridging process has obviously been slow and has affected small numbers of individual teachers, students, and institutional staff more than it has changed basic philosophies of instruction or learning on either side of the relationship.

A major challenge confronting further "cross-cultural communication" efforts is the need to foster broad-based awareness of the unique physical *and* intellectual resources offered by the informal educational institution. As one participant put it:

> The biggest hole [in current relationships between the science museum and formal education institutions] is that schools see the science museum as visitation resources, not intellectual resources.

Changing that perception may involve some focused advocacy by informal education institutions and the groups that support them. As one participant put it, the role of the science center may be to function as "an agent of propaganda" to legitimize play-like activity and, more generally, the experiential dimensions of science learning.

Curricular and instructional policies, often formalized in state testing and requirements, pose a second and related barrier. School people often have difficulty visualizing how informal science learning modes can help them meet these requirements. In some states, however, recent increases in requirements (e.g., for science instruction at the elementary school level) have brought educators to the door of the informal educational institution looking for help. Participants from New York State, for example, observed:

> In this state, since the Regents changed the elementary science requirements, there is a perception that the science centers can fix it. This is a wonderful opportunity. We can leverage it.

Whether or not the schools seek help from the informal education institutions, school curricula resist change, and for good reasons. For example, schools must protect themselves from special interest groups, each of which wishes to control some aspect of the learning process or its outcomes or both. Rather than seeking to change curricula, some of the institutions represented in the meeting had taken formal curricula as a given and tried to relate collection-based activities or exhibits to current instructional objectives. Doing so was not easy, however, and involved some obvious trade-offs.

Other significant barriers inhibit the formation of linkages between formal and informal education institutions, in particular, the following four:

- *Logistical problems.* Transportation to and from the informal education institution was frequently mentioned as a major constraint on developing better linkages. Some institutions had taken pains to organize transportation, but had found that raising funds to cover this kind of expense was particularly difficult.

- *Limitations on the physical capacity of informal institutions.* The physical plant of each informal education institution, no matter how large, is inherently limited in relation to the potential audience of school personnel and students. The more intensive forms of relationship (e.g., involving workshops for students or institutes for teachers) can be provided directly to only a small proportion of the population that the institution is hoping to reach. This fact has stimulated attempts by some institutions to achieve "multiplier effects," for example, by training a cadre of teachers who then go out and bring in others for similar training, but this only partially overcomes the constraints on space and facilities.

- *Limitations on the typical exposure time to the informal institution's resources.* As a corollary of the limitation on physical capacity, time for individuals to be in the institution's learning environment is typically limited--in most cases, visitors come only once unless the institution makes a special effort to encourage repeat visits.

- *School districts' lack of ability (or willingness) to commit resources to activities that enhance linkages.* Although the resources involved might be modest, some institutions have encountered reluctance on the part of school administrators to make resources available to support released time for teachers to attend training sessions at a science museum or transportation for students to and from the museum. In some cases, the lack of support from school administrators, combined with teachers' hostility toward these administrators, works against efforts to bring new ideas from the informal educational world into the schools.

Promising Entry Points

There are various "entry points" for establishing linkages with the schools that reflect differences in settings and the unique configurations of events, people, and opportunities confronting each institution. In part, the attractiveness of particular entry points depends on the informal institution's conception of its role in fostering a relationship with the schools. Four alternative conceptions of this role embody different philosophies of educational improvement and the institution's contribution to it:

- *Repository of unique intellectual and physical resources.* Almost by definition, each informal institution assembles unique physical resources, which draw people from the schools, but also provides a unique intellectual resource, including familiarity with the sciences and a philosophy of experience-based, voluntary learning.

- *Interface between basic research and the public.* In this conception, the informal institution acts as translator of scientific knowledge, including recent advances. Because it has high public credibility and relatively few constraints on the selection of content, the institution can present emerging scientific knowledge that is unlikely to be reflected in school curricula.

- *Safe haven for professional development and renewal.* The informal institution can offer a place for teachers and educators to develop as professionals independent of their particular work roles in the schools.

- *Change agent.* The informal institution can function as an active advocate for change in educational practices, curricula, and philosophies in local schools.

These conceptions of the institution's role vis-a-vis the schools alter the way opportunities for fostering the linkages are perceived. As change agent, for example, the informal science center is more likely to concentrate attention on existing curricula and teaching practices in the schools. As a safe haven for nurturing professional development, on the other hand, the institution's own facilities and collection-based programs are a more likely focus of its energies.

These conceptions of the informal institution's role are not mutually exclusive, however; most of the institutions represented at the meeting are attempting to assume more than one of the roles simultaneously.

Whatever the conception of role in fostering a relationship with the schools, mechanisms that increase school people's exposure to informal institutional resources build linkages that are particularly strong. Two ways of increasing the engagement of school people with institutional resources seem especially promising. Each emphasizes long-term relationships between the science museum and individual students or teachers:

- *Teachers and school personnel in museum "associate" roles.* In several instances, long-term connections with the schools have been fostered by retraining excellent teachers as museum-based workshop leaders, materials or exhibit developers, etc. These arrangements have a variety of side effects; for example, in some cases, museum exhibits have become more compatible with school curricula and outreach to other teachers has improved.

- *High school students in docent (or other) roles within the institution.* The use of high school students from communities targeted by the institution as high-priority audiences (e.g., inner-city ethnic neighborhoods) has apparently helped teachers and other museum goers "connect" with the museum exhibits. In addition, there have been effects on the docents themselves, some of whom are seriously considering science education careers as a result of their experiences as a docent.

Other possibilities have less to do with individual relationships. Informal institutions are also in an excellent position to play an intermediary role between universities and the schools, by bringing together the resources (both scientific and pedagogical) of the former and helping to translate these into terms that are useful to practicing educators. This has happened naturally at one of the informal institutions, an arboretum associated with a major university; but it can happen elsewhere in situations where the informal institution has no formal connection with a university.

Another promising entry point involves the potential of informal institutions to stimulate creative curriculum development. One participant put it this way:

> We're talking about the intellectual resources of informal education institutions. The people at these institutions are creative, strange, special thinkers. Teachers, on the other hand, are better adapters than creators. We ought to tap that resource.... Some of the best material development has drawn on work related to informal science education, for example, OBIS. How to tap this for basic curriculum development?

In support of this position, other participants mentioned the strong positive response they had seen from students and teachers to the sets of materials or kits that had already been developed through their projects. Students at one site were amazed that they were allowed to keep the diffraction gratings they had been given by science museum staff for experiments with light. The capacity to develop or identify appropriate materials for experiential science learning makes informal education institutions good at reducing this "starvation for materials."

Caveats in the Formation of Linkages

Although participants in the meeting were generally enthusiastic about the importance and possibility of forming linkages with the schools, they pointed out grounds for proceeding with caution. Three principal caveats emerged from the discussion: (1) certain important audiences (and schools) are difficult to reach and can easily be ignored; (2) informal institutions may compromise their unique strengths by focusing too much on school science programs; and (3) the staff of informal institutions may lack the expertise (e.g., in teacher education) implied by linkage activities.

Regarding the type of clientele informal education institutions serve, inner-city and disadvantaged populations--and the school systems serving them--are often hard to bring into long-term and meaningful relationships with the informal education institutions, although it is easy enough to attract individual students to museum exhibits and activities. In one instance, for example, a teacher support network serving multiple communities in a large urban area had found affluent suburban school districts more than willing to get teachers involved, while the inner-city district was reluctant. For many reasons, inner-city and disadvantaged populations are extremely important to reach, for they represent the segment of the nation's young people who are least well served by current science programs in schools and who face the prospect of advancing to adulthood without even minimal literacy in scientific and technological subjects. Without extra attention, effort, and, possibly, specialized strategies, these segments of the school population may not be reached effectively by linkage efforts.

Regarding the danger of focusing too much on the school science program itself, significant trade-offs exist when institutional activities are oriented to existing school science programs. The more closely museum exhibits or activities are tailored to existing curricula, the greater the risk of compromising the essential spirit of informal learning and discovery. Commenting on her institution's philosophy of educational programming, one participant noted:

> We have resisted the notion of going out into the schools. We feel that we can't do justice to the [wildlife] collection if we take it out of the facility designed for it...so we are increasingly bringing school into our collection. For example, in our "jungle lab," we built a classroom into a jungle in a sort of treehouse.

Other participants pointed out that most exhibits can't be easily translated directly into school programs. In exploring linkages with the schools, informal education institutions need to consider carefully where the "center of gravity" of their efforts lies--closer to the schools and their current curricula versus closer to the informal institution and its own program structure. In so doing, the informal institution must not compromise its unique strengths.

Regarding the actual capacity of informal education institutions to help school people, informal institutions may lack the necessary expertise in such matters as teacher education. Science museum staff are uniquely qualified to train teachers to use the institutional resources of the science museum and, by extension, to internalize the philosophy implied by museum resources. But they are not necessarily qualified to instruct teachers in pedagogy. This limitation in staff capacity reflects a more fundamental limitation alluded to earlier: by its nature, the learning environment within the informal educational institution cannot be replicated in a *formal* educational setting, except under exceptional circumstances (e.g., in some alternative schools). In this sense, informal institutions are unlikely to provide school people with an integrated and complete alternative to what they currently do in school science programs.

Implications for the Foundation

Discussion throughout the day remained primarily at the level of the project, the problems it addressed, and the solutions that might be found. But the ultimate purpose of discussion was to generate ideas for NSF as it considers future investments in this area. By interpreting participants' remarks, we can shed light on four questions regarding the Foundation's current approach to investing in this area. Should NSF:

(1) Target its investment efforts in this area toward broad goals or specific programmatic entry points?

(2) Alter its current array of programs to address this investment area more directly or effectively?

(3) Offer different forms of support (e.g., longer-term funding, funding for currently disallowed purposes)?

(4) Adopt a more explicit "advocacy" stance to encourage the development of linkages between schools and informal educational institutions?

We preface our interpretations with the observation that participants from the field had understandable difficulty addressing questions of Foundation funding strategy. These participants saw the issues through the lens of their own local settings and their institutional positions. From these vantage points, it is not easy to visualize how a public foundation at the federal level should direct its funding to further a broad investment goal.

Nature of NSF Investment Targets

There appeared to be general agreement that rather than targeting specific types of entry points (e.g., teacher associate roles, traveling-kit design), NSF was better off establishing a broad and strongly stated goal of fostering linkages between the informal institutions and the schools. The programmatic specifics would then be left to the imaginations and unique circumstances of proposers. But directing funds toward other related investment targets can enhance the chances that investments in linkages with the schools would pay off. By the design of its funding solicitations or its outreach to potential proposers, NSF might:

- Promote the idea of the informal institution as a neutral player that could bring together school, university, and other resources in efforts to improve science education.
- Nurture linkages among informal institutions themselves to encourage the sharing of information and ideas related to this area.
- Address NSF funding in this area more explicitly to the needs of underrepresented groups.
- Target research support for investigations into topics related to the linkage formation process: what attracts school teachers or students to informal education institutions, how students (or teachers) "play" with materials and what kind of experience they have in this situation, and what they take away from these experiences (see discussion in Section V of assessment approaches relevant to this goal).

Adequacy of the Current Program Structure

Some or all of the targets for investment just described could become part of a strategy designed to foster linkages more purposefully than at present. But one needs to ask whether such a strategy will or should happen under NSF's current program structure.

Most promising activities for establishing or improving linkages between informal science education institutions and the schools can be supported under existing NSF programs. Although none of the six programs represented at the meeting currently makes this investment area a priority, the possibility of support for a wide range of linkage projects is there, as suggested by Table IV-2. Given this fact, it is probably unwise to consider radical alterations in existing programs; none were proposed during the meeting.

One clear exception was noted: current NSF programs (with the exception of the Young Scholars Program) do not support the teaching of children directly, nor do they permit direct support to students in any form. Although there is a rationale for

Table IV-2

HOW CURRENT SEE PROGRAMS CAN SUPPORT INVESTMENTS THAT ESTABLISH LINKAGES BETWEEN INFORMAL SCIENCE EDUCATION INSTITUTIONS AND THE SCHOOLS

SEE Programs and Initiatives	Types of activities in this investment area that each program supports
Informal Science Education	▪ Low-unit-cost, direct-impact projects aimed at young people and the public, which can feature linkages to the schools.
Instructional Materials Development	▪ Innovative or updated K-12 curricula that are aimed at widespread distribution, e.g., through commercial publishers--these can be developed by an informal institution and/or can emphasize informal learning modes.
Teacher Enhancement	▪ Teacher inservice training, especially at the elementary and middle school levels--informal institutions are eligible so long as they meet a series of criteria for good inservice training projects.
Teacher Preparation	▪ Innovative approaches to teacher preparation--these could involve greater ties between the university, the informal institution, and the schools.
Science and Mathematics	▪ Dissemination, conferencing, and the Education Networks development of networks--informal institutions can be part of these networks or a central focus of conferences or network investments.
Private Sector Partnerships to Improve K-12 Science and Mathematics Education	▪ Collaborative projects of all kinds that feature the private sector as an intellectual partner--informal education institutions and the schools can be a part, or even the central players, in these projects.

this policy--that NSF is in the business of catalyzing improvements and building institutional capacity--one can also make the argument that direct support to certain categories of teaching in informal settings or even to students would do much to build the desired linkages. For example, explainers in science museums might receive a kind of "associateship" award to support and motivate their involvement with the sciences and science education over the long term. (For logistical reasons, NSF would need to find an organizational middleman, such as the Association of Science and Technology Centers, to dispense these awards, a role analogous to the National Science Teachers Association in the Presidential Awards for Excellence in Science and Mathematics Teaching.) As alternative forms of programmatic investment are considered in the Directorate's strategic planning process, this category of support deserves to be included in discussion.

But aside from this possibility, the important question is this: Is the Foundation likely to receive the right mix and quality of proposals aimed at this investment area, given NSF's current program solicitations and outreach to the professional community? Although participants in the meeting did not address this question directly, we draw the conclusion that relatively few proposals are likely to arrive that take the establishment of linkages as a central goal. It is more likely that proposals to the Teacher Enhancement program, for example, will tend to emphasize its priorities and will focus primarily on continuing-education goals; proposals to the Instructional Materials Development Program will aim at curricular innovations that may or may not feature the approach or materials of informal science institutions. Some good linkage proposals will probably be funded, but the goal of establishing linkages will remain low in priority, a by-product rather than a primary aim of funded projects.

If this investment area is to be addressed more explicitly and extensively, SEE may want to consider one or more of the following actions:

- *Conducting more aggressive outreach to relevant segments of the professional community.* Including meetings such as this one, SEE could do much more to cultivate the network of groups and individuals who are or might be interested in undertaking projects in this investment area.

- *Altering the priority statements of current program announcements when they are next revised.* By including linkage goals as a priority in revised program announcements, SEE could signal more directly its interest in attracting proposals aimed at establishing linkages between informal education institutions and the schools.

- *Adjusting the review process to give special attention to this investment area.* Instead of reviewing proposals for linkage projects alongside other proposals in each of the respective programs, a special panel might be set up to review these proposals in relationship to each other, despite the fact that they might propose different kinds of activities (e.g., teacher support activities, materials development).

- *A cross-program initiative aimed at this investment area.* SEE might also wish to issue a cross-program initiative announcement, analogous to *Private Sector Partnerships to Improve K-12 Science and Mathematics Education*. This initiative could be set up with its own budget, but be designed to fund proposals jointly with any SEE program that is relevant to a particular proposal effort.

How important is it for NSF to support the goal of establishing these linkages, as opposed to other worthy goals funded by the respective programs? The meeting provides no basis for answering the question, but the question should be asked--and may not have been asked to date. This investment target may deserve a higher priority in some or all of NSF's programs and in its overall strategic planning for support of science education. Although they have clear limitations, informal education institutions are an underused resource in efforts to improve science education in the schools. Because these institutions can engage young people (and their teachers) in science learning and contribute to general scientific literacy, they have a great deal--in principle--to offer the schools.

Change in Funding Requirements To Facilitate Investment in This Area

NSF currently provides grants based on a catalytic funding philosophy: the Foundation assumes that by granting funds for several years (up to 5 years in a few cases), important innovations will be initiated, which will attract other sources of support over the long term. Participants in the meeting considered the possibility that longer-term funding (e.g., for 5 to 10 years) was necessary to support the long-term formation of linkages with the schools, but were generally agreed that this wasn't critical to the success of a local linkage initiative. In most instances, participants' projects had built relationships with the schools using Foundation funding as one (often the first) among many sources of support. The lack of continuing NSF support was not viewed as a big obstacle to achieving project goals. Some participants also raised the possibility that long-term support might become viewed as an "entitlement," similar to formula or block grant funding, thus creating a disincentive for competitive performance, which lies at the heart of NSF's grantmaking process.

Other, smaller alterations in NSF's funding requirements were also discussed, such as the desirability of permitting NSF funds to be used for transportation costs. Although the rationale for disallowing this category of expense is clear enough, there are reasons why making exceptions to the policy might make sense in this investment area, because the physical movement of people is so central to the nature of linkage activities.

The Possibility of an Advocacy Role for the Foundation

Whatever else it does to promote investment in this area, NSF has the option to adopt a more visible posture in promoting linkages between informal education institutions and the schools. This will happen anyway if the Foundation makes this investment area a higher priority and so indicates in its program announcements or outreach to the professional communities, as discussed above. But NSF can assume an advocacy role apart from what it does to attract and fund proposals in this area. NSF can try to project one or more visions of the relationship between schools and informal institutions as a way of orienting relevant actors toward possible actions in this area.

Various activities might contribute to this kind of advocacy function. At the least, NSF might consider:

- Preparing a position statement (independent of program solicitation announcements) as a desirable focus for improvement efforts.

- Inviting knowledgeable individuals to prepare commissioned papers on the problems and prospects for fostering linkages between informal education institutions and the schools.

- Hosting a symposium or conference to put together key actors representing all sides of the linkage relationship to explore the issue and develop visibility for it.

- Interacting with relevant associations--the Association of Science-Technology Centers and the American Association of School Administrators, for example--to stimulate dialogue on the matter.

The advocacy role implied by most of these mechanisms is indirect, unlike the more direct "bully pulpit" advocacy that might be less appropriate to NSF. The Foundation must always be wary of being in the position of prescribing educational solutions from the federal level. But indirect advocacy may nonetheless be powerful and, to the extent NSF sees this as a high-priority area, it should count these mechanisms among its important tools.

But the advocacy role takes on greater importance when one realizes that an initial barrier to the formation of linkages between informal education and the schools lies in the recognition of the concept itself. Those who attended our meeting are among the few that have begun to recognize the importance and possibilities of establishing a long-term relationship with the schools. In one way or another, they expressed the belief that connecting with the schools is somehow *fundamental* to their educational mission and role in the community. Others--perhaps the majority of school-based educators and informal institution staff--have yet to take this view or even consider it as a possibility. In this regard, the Foundation may play an important role in raising this issue to the level of awareness and possibility it deserves.

V AN EXPERT MINI-CONFERENCE: EXPLORING THE ASSESSMENT OF LEARNING IN INFORMAL SCIENCE SETTINGS

By

Mark St. John
Patrick M. Shields

V AN EXPERT MINI-CONFERENCE: EXPLORING THE ASSESSMENT OF LEARNING IN INFORMAL SCIENCE SETTINGS

This pilot activity--a mini-conference of NSF staff and experienced informal science educators--was aimed at exploring the issue of how one might study NSF's investments at a "micro" level: in particular, how individuals interact with NSF-funded informal science education resources and what they "learn" from those interactions. The issue of assessing individual learning is an important one for NSF because the justification of its investments ultimately hangs on the current belief that people gain something educationally valuable from their interaction with informal education resources.

Highlights

There were several major themes that recurred throughout the discussions of the 2-day meeting. We have grouped these themes into six general "guidelines" for approaching the assessment of informal science learning. It is important to remember that these guidelines are not "findings" derived from hard data, but rather the collective wisdom and opinion of the meeting's participants.

(1) Informal science education should not be thought of as a set of "learning" experiences similar to formal schooling. The goal of informal science education is not primarily the teaching of specific content and skills.

(2) Rather, the central mission of informal science education is broader. Informal science education resources can provide an interface between mainstream culture and the subculture of science, mathematics, and technology. Thus, informal science education is best thought of as helping to "acculturate" individuals to the scientific world.

(3) Assessments in the informal domain should explore and document how informal science education resources contribute to the individual's (and the nation's) acculturation process. To assess how informal science experiences affect individuals requires a very broadly defined set of outcomes and a long-term view.

(4) Assessment of informal science learning at the project level should be "formative" in nature--that is, designed to provide feedback that helps projects design resources that are appropriate to the knowledge, interests, and attitudes of the audiences they are meant to serve.

(5) Assessment of informal science learning at the initiative level should be prospective and be designed to inform NSF about the assumptions underlying its initiatives. Because answering NSF's questions requires more than the

aggregation of project-level assessments, cross-project studies and studies of critical projects should include multiple, complementary assessment methods that in combination generate a "mosaic" of evidence about the way these investments affect individuals.

(6) Overall, a program of applied research is needed to help articulate the goals and mission of informal science education as well as to search for new ways to think about and approach the assessment of informal science learning.

Design of the Meeting

For several reasons, we believed the topic of assessing informal learning to be a good candidate for review by a small working conference of experts. Unlike other pilot activities described in this report, we did not attempt to carry out an illustrative pilot study of informal learning. In doing small studies to describe large areas of investment or suggest the overall effects of an initiative (see Sections I through IV), we were able to draw on common assessment procedures (interviews, site visits, document review, secondary analysis) with confidence that our findings would be both useful and appropriate to the questions posed. In contrast, when examining informal science learning, the main issues for assessment are more conceptual than technical and require a rethinking of informal learning itself and NSF's role in supporting it. Consequently, we felt the greatest contribution we could make would be to "back up a step," seek broader perspectives, and reconsider the question of assessing informal science learning at the individual level.

Our working session was designed to provide an opportunity for NSF staff to explore this difficult and important assessment issue with the best minds in the field. NSF program and division officers, caught up in the daily pressures of processing proposals, rarely have the chance to spend a day or two exploring fundamental questions of Foundation strategy or policy, especially in the area of assessment. Even more rarely do they find the opportunity to involve a handpicked group of experts in their deliberations. Thus, our working seminar sought to illustrate a mechanism by which SEE program and division officers could engage in a process of reflecting on large, long-term issues.

For this meeting, we brought together experts deliberately chosen to represent diverse fields and perspectives on assessment. The list of invited participants appears in Table V-1.

To provide structure for the meeting, as well as to introduce a common framework for the discussion, we prepared a paper and distributed it to all participants before the meeting. This discussion paper (which is reproduced as Section VI of this volume) suggested three broad questions as the focus for the discussion:

Table V-1

PARTICIPANTS IN EXPERT MINI-CONFERENCE ON THE ASSESSMENT OF INFORMAL SCIENCE LEARNING

Experts on Informal Science Education, Research Assessment

Valerie Crane (Research Communications, Ltd.)	Communications and marketing researcher
Elsa Feher (Reuben Fleet Space Theater and Science Center)	Physicist, specialist in cognitive studies in science museums
Keith Mielke (Children's Television Workshop)	Research specialist in children's science television
Roger S. Miles (British Museum of Natural History)	Exhibit designer and researcher
Jon Miller (Northern Illinois University)	Political scientist, specialist in scientific literacy
Philip Morrison (Massachusetts Institute of Technology)	Physicist, science educator, author
Phyllis Morrison	Elementary science specialist, film and book reviewer
Vito Perrone (Carnegie Foundation for the Advancement of Teaching)	Evaluator, expert in inquiry-based science learning
Bret Waller (J. Paul Getty Museum)	Artist, art historian, art educator

National Science Foundation Staff

Merlyn Behr	Research in Teaching and Learning Program
Ray Hannapel	Research in Teaching and Learning Program
Michael Templeton	Informal Science Education Program
George Tressel	Division of Materials Development, Research, and Informal Science Education

(1) What kinds of "learning" are most important in informal science education? In posing this question, the paper outlined in a schematic way the logic of NSF's informal science investments and how they might help enhance the overall "acculturation" of individuals into the world of science and technology. This idea of acculturation was then related to the most commonly articulated goals of informal science education.

(2) What approaches and procedures could be used in assessing these outcomes? In discussing this question, the paper reviewed past and existing approaches to studying informal science learning.

(3) Which of these possible assessment procedures should receive highest priority for NSF funding? In raising the question of priorities, the paper discussed other factors to be kept in mind--such as the different audiences for assessment information and the differences between media.

At the meeting itself, SRI staff served as facilitators, moderating the discussion and keeping it focused on the issues central to NSF. The facilitators played an active role in framing the discussion, both offering broad issues to address and focusing on specific questions around which there is considerable debate (e.g., is it important for informal science resources to address current prominent issues of science and society?). Thus, the meeting was not unlike an extended focal group discussion with the ultimate aim of exploring a range of perspectives on the proper approaches to assessment in this area of investment. The meeting lasted 1-1/2 days, with each 1/2-day session addressing one of the questions above.

Guidelines for Assessing Informal Science Education Learning

The themes and issues that emerged from the analysis of past assessment work and from the discussions that took place are presented in several different formats:

- *A short summary*: The major themes and recommendations of the conference, as well as a discussion of the relative merits of this type of working group meeting, are included in Volume 1 (see Section IX).

- *A synthesis of themes*: In this section of Volume 2 we have synthesized and summarized all that we have learned in reviewing the assessment literature and the discussion of the working group meeting.

- *A reconstructed dialogue*: The appendix to this volume is intended to provide the reader with a more direct and rich account of the working group discussion, in the form of a reconstructed dialogue.

The discussion below lays out six general guidelines for shaping future efforts by the Foundation to assess what individuals learn from informal science education experiences.

Informal Science Education Should Not Be Thought of as a Set of Learning Experiences Similar to Formal Schooling

Determining assessment methods or even basic philosophies of assessment that are both appropriate and useful in this domain has proven to be a very difficult task, largely because the nature and consequences of informal learning experiences are not well understood. As a consequence, the practitioners of informal science education are extremely skeptical about assessment approaches that derive from the formal domain and a "testing mentality."

Many articles have been written on the unique nature of the informal learning environment, yet the almost unconscious tendency of evaluators working in this domain is to persist in trying to understand the purpose, the activities, and the outcomes of informal learning in terms of the concepts derived from formal education. Nevertheless, several aspects of the informal science learning enterprise and environment differentiate it from the kind of science learning that takes place in the schools. The following five differences are especially important.

First, the audiences for informal education are extremely diverse. The age, background, science interests, and intellectual sophistication of the people who comprise these audiences reflect the entire spectrum found in the population. Consequently, one cannot think about designing instruction for an audience as one might do when writing a high school physics text.

Second, informal science education occurs in a context of recreation and entertainment, with characteristics specific to the media and arena in which it occurs. Children watch "3-2-1 Contact!" for much the same reasons as they watch noneducational programs; people go to a science museum on Saturday in much the same way that they go on any other Saturday outing. Informal science educators must design educational resources that remain faithful to the nature of the recreational context in which people will encounter the resources--otherwise they risk "preaching to people in empty halls."

Third, in the informal learning setting the audience is in control of the entire interaction. Instructional designers and informal science educators can have only little influence over what those in their audience do, the order in which they do it, what they attend to, and what (if anything) they are interested in learning about. Informal science resources are primarily cultural and recreational resources; only secondarily do they serve as educational resources. To think of them primarily as instructional media will miss the essence of their character and misconstrue the value they have for most people.

Fourth, in general, the "science" that is included in informal science education is broader than the "science" included in the school courses. There are, for example, many science-like activities in the informal domain (birdwatching, model trains, sailing, computer games) that may have important mathematical or scientific educational value for young people.

Fifth, in addition to addressing a wider range of topics than schools, informal science education promotes a wider range of learning modes--especially nonverbal (iconic, haptic, kinesthetic, tactile) modes of learning. Conveying verbal knowledge should not be taken as the primary goal of informal science education; this is probably better left to the schools.

The Central Mission of Informal Science Education Is "Acculturation" to the Scientific World, Not the Teaching of Specific Content and Skills

By drawing an analogy with the prominence of sports in our world, one participant pointed out how science, mathematics, and technology comprise a large subculture within our society:

> I just wanted to say that I very much like the notion of "acculturation" as a goal for science education. This helps us to broaden our thinking about what we are doing. It is true that science and technology comprise a major subculture, and are increasingly evident in our mainstream culture. It is important that more than just scientists and engineers be acculturated.... A second, much larger tier of the populace must be able to think scientifically as well.... And perhaps it is only the subculture of sports that is equally, or even more, pervasive than that of science. It might be very interesting for NSF to study sports in order to better understand how people become a part of a subculture. After all, scientists, like professional athletes, ultimately need a knowledgeable and appreciative audience to play to, and in sports there is such an audience. Kids grow up playing sports; sports are discussed in the family; they are on television; there are sports sections in the paper.... A multitude of informal processes throughout life combine to generate a "sports literate" population.... I think this is a useful parallel for thinking about the development of science literacy.

As we assert in our discussion paper, to become scientifically literate means to become more familiar with, and more a part of, the "subculture of science, mathematics, and technology." What is involved in being part of a culture? In an anthropological sense, when people grow up in a culture, they develop a personal knowledge of the ideas, vocabulary, issues, important people, institutions, and history of that culture, their view of the world is shaped by attitudes and perspectives that derive from that culture, and they have intellectual capacities that are selectively developed by the educational priorities of that culture. Similarly, acculturation into the world of science includes gaining knowledge, developing attitudes, and increasing intellectual capacity. These aspects of acculturation include many of the often-espoused goals of science education, as described in Table V-2 below.

Table V-2

GOALS OF INFORMAL SCIENCE LEARNING, AS OUTCOMES OF AN ACCULTURATION PROCESS

Transmitting Knowledge

- Cultural scientific literacy
- Civic scientific literacy
- Knowledge of the scientific enterprise

Fostering Positive Attitudes

- Excitement about science
- Motivation for further learning
- Improved attitudes about the scientific enterprise
- Changed sense of self in relation to science

Building Capacity for Further Learning

- Experimental scientific literacy
- Changes in misconceptions
- Scientific habits of mind

Informal science resources provide an interface--a doorway--between the mainstream culture and the large subculture of science, mathematics, and technology. These resources (museums, television, magazines, etc.) allow those who work in science to present and translate the ideas and perspectives of their world to individuals in the mainstream culture and, in return, provide an opportunity for these individuals to share the world of scientific discovery. In this way, by allowing the different cultures to infiltrate each other, informal science education investments can contribute to the acculturation of individuals (and the nation) to the world of science, mathematics, and technology.

A well-known science educator expressed the idea in this way:

> We live in a scientifically and technologically driven economy, and we live in a culture that throughout this century has been identified with science--the atomic age, the space age, the computer age, etc.... The use of total resources (formal and informal) for science education has become a question of education for living in a science and technology culture. This makes science education a matter of acculturation--so far a little-recognized goal.... Formal and informal education are ways (properly conceived) of making it possible for children not to be foreigners in their own culture.... For these reasons, I cannot be unbiased about the importance of informal education in the sciences.... (Hurd, 1986)

There are several advantages in using the idea of "acculturation in the sciences" as an overarching goal for informal science education. This concept:

- Focuses attention on long-term knowledge or attitudinal "gains." Acculturation implies repeated cumulative experiences with science in both formal and informal settings for individuals to internalize the culture of the world of science and technology.

- Implies a great deal of interactive experience. In fact, the degree to which an individual feels acculturated in mathematics and science is the joint result of many different experiences and influences, including the kinds of resources NSF funds, other informal resources (e.g., newspapers, television news), the home environment, the workplace, and more formal learning experiences in school.

- Connotes a lifelong process of developing scientific interests, knowledge, and habits of thought throughout one's life.

- Is highly compatible with NSF's overall mission of helping to broaden the pool of people who are competent and interested in science.

School learning typically focuses on one aspect of the acculturation process--on the transmission of knowledge and the development of a cultural and civic scientific literacy. Informal learning experiences are probably best at promoting other, more ineffable aspects of acculturation. For example, many scientists point to important informal learning experiences that excited them about science and helped them to sustain and further develop their interests in science, as noted by research on the educational benefits of the apprenticeship role that most Nobel Laureates served early in their careers:

> One point on which the Laureates are largely agreed is that the least important aspect of their apprenticeship was the acquiring of substantive knowledge from their master...the Laureates testified that the principal benefit of apprenticeship was a wider orientation that included standards of work and modes of thought. They report, in effect, that the apprenticeship was a time of what social scientists call socialization.... Socialization includes more than is ordinarily understood by education or by training: it involves acquiring the norms and standards, the values and attitudes, as well as the knowledge, skills, and behavior patterns associated with particular statuses and roles. It is, in short, the process through which people are inducted into a culture, or subculture.... (Zuckerman, 1977)

Two Nobel Laureates describe their own acculturation as apprentices in the following way:

> It is the contact: seeing how [masters] operate, how they act, how they go about things.... It is not at all the specific knowledge. It's learning a style of thinking, I guess. Certainly not the specific knowledge...there were always people around who knew more.... It wasn't that. It was a method of work that really got things done....
>
> I knew the techniques of research. I know a lot of physics. I had the words, the libretto, but not quite the music. In other words, I had not been in contact with men who were deeply imbedded in the tradition of physics...this was my first real contact with first-rate creative minds at the high point of their power.... (Zuckerman, 1977)

The importance of conveying the real flavor of the discipline appears to be essential in terms of engendering deep and long-lasting impacts on personal interests and career choices. Thus, by funding high-quality informal learning resources, NSF hopes that its investments will present the enterprise of science in such a way that it stirs the emotions and changes the attitudes of individuals of all ages--outcomes that lie at the heart of the acculturation process. Exciting experiences with real phenomena may be able to provide motivation for further learning about a topic. Well-designed informal science resources may provoke curiosity as well as build confidence so that the next encounter with the same topic stirs further personal interest and further learning about the topic. In short, positive experiences with informal science resources can initiate a cycle of curiosity, approach, learning, confidence, and renewed curiosity. With repeated experiences, a deeper sense of confidence may build to the point that one's sense of relationship with science changes, even to the point where one begins to become a "lover" or "connoisseur" of science and to identify with it.

Informal science education may also have the potential for building capacity for later learning, which furthers the acculturation process, as well as supporting formal learning. Encounters with informal science resources may help to build a base of direct experience and knowledge of the objects and phenomena of science. One participant in the conference referred to this as learning about the substance as well as the symbols of science. This kind of learning, some argued, provides for an internal matrix of experience that facilitates later, more formal, conceptual learning.

Assessment Should Explore and Document How Informal Science Education Resources Contribute to the Individual's Acculturation Process

There are real difficulties in assessing the kinds of ineffable outcomes that are part of the acculturation process. Some meeting participants argued that assessment of these outcomes at the individual level should not be done at all, because the nature of informal learning could be so easily distorted by subjecting it to inappropriate tests and measures. It is clear that, to be valid, assessment approaches must respect the diverse nature of informal learning.

For example, "input-output" methods of assessing what people learn generally do not work well in informal education. This approach, commonly used by evaluators in formal education settings, rests on assumptions about the relationship between instructor and pupil, which don't pertain to the informal education settings. Thus, when evaluators attempt to measure the extent to which the goals of the television or museum designer are achieved, they almost invariably come to the conclusion that very little learning is taking place. The question that remains unanswered is whether the failure lies with the project, or whether it more accurately represents the failure of the evaluator to look broadly enough, in the right places, or with instruments sensitive enough to see all the different ways that the experiences may contribute to "learning." As one participant noted:

> The danger in [input-output] models of evaluation comes not from the fact that they are rigorous, but more in the assumptions they are making about the fact that you can capture informal learning using a very structured way of measuring the phenomenon. I mean they are measuring by saying, "Here's a set of knowledge and by the time you go out the door you should have it." So it's the assumptions of the people doing the research that are off base.... They are making value-laden decisions about expectations and about the fact that the learning experience has to be sequential, cumulative, and equal for everybody.

Given the complexity of the informal science learning experience, it makes more sense to focus assessment initially on the task of describing what is happening in informal learning settings. Documentation--both statistical and qualitative--should thus play a large role in all evaluations of informal science learning. One scholar at the meeting put it this way:

> I think we should get more into the processes of learning and move away from the boxes and the grades and into understanding more about how informal learning works. We have not spent much time talking about that here. We have to look at it much more in detail and in context. We have to understand what kinds of individuals come into the loop, what happens to them once they get there, what this experience is like for them--really tracking in a very careful way what is this process we are talking about, trying as we go along to define it. I think it's much more a process of looking at the details of what happened. We know lots and lots of things go into each person's experience, and I don't think we should worry about giving grades.

This implies that assessment centered either on the broad goals of NSF or the behavioral objectives of the exhibit designer can miss much of what is happening and overlook much of what people are learning. Rather, some participants argued that evaluation should give high priority to the "consumer's point of view" and not be blinded by the perspective of the educational designers. Just as *Consumer Reports* ignores the claims of the manufacturers and only looks at products along dimensions that are important to the consumer (safety, reliability, cost-effectiveness, flexibility, etc.), so evaluation of informal learning should be largely "goal free" or "audience centered." In this way, the perspective of the educator and the designer ultimately can be brought into line with that of the consumer. One designer succinctly described his experience with such an approach:

> We did this at the children's museum, and it was terribly useful to us, right there and then. I would simply stop kids as they left the museum and say "Tell me about what you have been doing." And essentially what I was listening for were things I hadn't heard before, things that I didn't know about--I was looking for things that were surprising to me. And I got a lot. You have to ask and then listen very hard, and when you do that, you hear some very surprising things, and those surprises were very good to hear.

Another implication is that the value of NSF's informal resources should not be judged solely or perhaps even primarily on the basis of the empirical evidence that people acquire knowledge from them. It may be a mistake to assess informal education resources from the point of view that they are the main source of learning about a phenomenon or even the main determinant of one's attitudes about science and mathematics. Much of their impact may come through complicated and subtle interactions with many other sources of information. Informal learning resources may thus contribute to acculturation without having a single or "main" kind of impact. As one participant asked, "Do you get credit for contributing to a wide range of interactions that ultimately affect the person, or do you have to do the whole job yourself before you get any credit?"

Moreover, television and museums are cultural resources. They are used by people for social and recreational purposes. They provide a cultural opportunity and an easily accessible option to interact with aspects of science and technology. It may be no more appropriate to judge the value of this cultural resource through the measurement of learning than it would be to measure the value of the opera or symphony by pre- and posttesting its audiences.

Finally, assessment should aim at learning lessons for the future. It should help build knowledge about the nature of the enterprise. Good projects do not only have good news to tell. The most persuasive and useful assessments are those that give an honest rendering of the good news and the bad news. The intent of the evaluation enterprise, given a realistic appreciation of what it can and cannot measure, should not be to "give report cards and grades" to the projects that are assessed. For these reasons, most evaluations should be designed to serve "formative," rather than "summative" purposes. From this perspective, the most useful goal of evaluation done at the end of a project is to inform future efforts. One participant with years of experience in evaluating informal science resources said:

> I don't like the words "formative" and "summative"...summative evaluation usually means "tell me whether this was a good project." That doesn't mean anything to me. We're trying to find out how to make something happen, and in that context, there is no "bad" project. If it doesn't work, we want to know it and we want to know why so we won't do it again or so we can figure out what else to do--it's all formative evaluation from our point of view; it's future oriented.... It tells us the good news and the bad news so that we are in a better position for the next stage.

The field now requires a broader and clearer rationale for the investment of public money in informal science learning. It is premature to talk of the effectiveness or the cost-effectiveness of investments in this area, since the overall goal of the enterprise at the individual level is not well understood. The attempt to describe informal learning as part of acculturation is but one example of the kind of rationale-building that is needed. Assessment efforts must help in building this rationale, first, by articulating broader visions of the enterprise and, second, by developing evidence that the assumptions underlying the rationale are sound.

Most Project-Level Assessment of Informal Science Learning Should be "Formative" in Nature

Project-level evaluation should concentrate on assessments that can serve a formative function for the project itself--that is, help the project team improve what it is doing in midstream. The main need for such assessment is to help create congruence between the learners' (visitors', viewers', etc.) knowledge, attitudes, interests, and motivations and the designers' goals, intended messages, and hoped-for impacts. There is no sense in doing extensive "summative" evaluations on projects that were misconstrued from the beginning. One meeting participant said, "Out of every $10 you spend on assessment, you probably ought to spend $7 or $8 before you are through, and leave a dollar or two to review things at the end."

Formative evaluation is part of an iterative process of prototype design and testing, with feedback being gathered from audiences at each stage. By integrating audience feedback into considerations of change in design, the formative evaluation process allows exhibits, television shows, or other informal learning resources to evolve as they are constructed--which is important in an area where good design is difficult to achieve a priori by even the most skilled practitioners.

Useful techniques include:

- *Naturalistic observation*--Close, detailed, unobtrusive observation of how visitors actually use the resource can help designers infer the kinds of educational outcomes that are likely (or not likely) to result.

- *Cued testing or "facilitated interviewing"*--By mediating the audience's interaction with the resource, and through detailed personal interviews, the designer can find out what people can learn from the resource under the best possible learning conditions. This semi-experimental "best case" approach sets a kind of "upper learning limit"--an asymptote that actual use in an informal learning environment may approach.

- *Entry interviews*--These can help designers understand what levels of knowledge, interests, and motivations people bring to their interaction with an informal learning resource. Far too often designers make implicit assumptions about their audiences, with the result that the resources they design fail to offer what the audiences are interested in or what they can understand.

- *Exit interviews*--Done in an open-ended and audience-centered fashion, these interviews can explore what people remember, found to be surprising, and liked (or didn't like). The exiting audience can even be asked to critique different parts of the project design.

- *Market research*--Using interviews or focal groups, the techniques of market research can help planners and designers very early in the planning process choose among topics and design schemes. Designs can then be guided by data regarding what is most likely to interest and engage audiences.

Several other points about formative evaluation at the project level need to be kept in mind. First, predetermining and then measuring the achievement of a project's "learning objectives" (especially when stated in behavioral terms) has limited utility for learning what is happening or communicating to designers how to do their job better. One participant summarized the point in the following way:

> I think it is very important to remember that one of the defining characteristics of informal education is the enormous variance in the audience that comes to it in the first place. When we talk about specifying and measuring behavioral objectives, that is coming out of a tradition of homogeneity--where you can say, "For this slice of instruction, with these assumptions, I would

> like this message to have this effect." But it's extraordinarily difficult to take something that is open to the full sweep of television--from the very young preschooler to the very old person watching the same program--to say that it makes any sense at all to talk about a behavioral objective guiding that enterprise...
>
> I think there is almost a disservice that's been done by this kind of up-front criteria setting because...not only does the viewer or visitor not behave that way, the instructional designer doesn't behave that way either...they're totally incapable of designing an exhibit based on the idea that 80% of the people should push a lever. That's not the way the design works. Maybe that is why there has been a lot of resistance to incorporating the results of museum research--because some of the research models that are out there are so off-base....
>
> You just can't specify exactly the nature of the transaction with the exhibit... [I find this approach] incompatible with the nature of the informal environment....

Second, the design of informal science education projects is an art that requires a general understanding of the basic nature and principles of communication. Although there was some debate about the point, the participants felt that assessment and/or research studies were more likely to help answer specific questions in specific set tings than to yield well-defined design principles for all of informal science education. In short, informal science education is likely to remain more of an art than a science for some time (and it should be assessed from this point of view as well). It appears that what makes for a good exhibit or television show is no more likely to be pinned down by research than are the style and characteristics of good teaching:

> One might, in principle, conceptualize a grid of objectives consisting of a very complex set of input characteristics, demographic clusters, and an equally complex set of outcomes.... When I was in graduate school, I really did believe that if we just could identify and control enough variables, we'd finally get them all into the corral....

Third, the evaluation expertise needed to do good formative evaluation at the project level tends to be limited, if not largely absent. NSF might encourage the development of increased expertise among practitioners by:

- Insisting (to a reasonable extent) that projects include formative assessment procedures in their designs.
- Helping to build the capacity of the field to do such evaluations by enabling projects to share their evaluation experiences and knowledge. (This includes sharing not only "successes" but also the "horror stories" of projects gone wrong because of the lack of forethought and assessment.)

- Funding small efforts to do training in this area for in-house staff.

- Supporting attempts by informal science practitioners to share evaluation resources (the Exhibit Research Collaborative is an example of how such resources can be shared--see Section I).

Assessment at the Initiative Level Requires a "Mosaic" of Approaches

Assessment that helps NSF learn about its initiatives is different from assessment designed to help those working on informal science education projects achieve their own purposes. Put another way, project evaluations cannot be aggregated to provide a full assessment of the initiative that spawned the projects. For example, project members might want to answer the question, "Does this exhibit do what we hope it will?" At NSF's level it may be sufficient to answer the more general question, "Do good interactive science exhibits have a positive influence on people's knowledge, attitudes, or behavior?" Thus, for NSF to evaluate its initiatives in this domain, special assessment efforts will need to be funded over and above project evaluations. It is probably best that these efforts be funded as separate activities (e.g., carried out by third parties or through other mechanisms described in Volume 1). When assessment is left as a "desirable feature," it is unlikely to receive the concentrated attention it deserves.

To learn about informal science education initiatives, it is better to focus on a few critical or exemplary projects (e.g., "3-2-1 Contact!") than to try to measure learning in all informal education projects. In doing these studies, multiple methods and perspectives should be combined to provide as comprehensive a picture as possible. In assessing the complex process of learning in informal settings, different methodologies can complement each other. One participant advanced the notion of a "mosaic" of studies:

> You have to remember there are weaknesses, large weaknesses in all approaches.... So that is why I like the concept of a mosaic...of studies, where the strength of one design compensates for the weaknesses of another. Collectively, then, we have an array of studies that interlock. Within that array, a true experiment can be a very powerful addition.

This participant went on to provide an example of how a mosaic of methods can be used effectively:

> I like the concept of an interlocking mosaic design...I know that certain effects are not going to be visible at all in some of the designs that are akin to taking a snapshot from a satellite a million miles out in space. You get the big picture but you miss an awful lot. On the other hand, I am reminded of a study we did of a single segment of a TV show in which we tried to track, cognition by cognition, whether a certain problem was understood, whether the underlying mathematical principles were understood.... That was laborious data

> **collection that could not be touched with a 10-foot pole by other methodologies of longer term and broader scope. There is always a trade-off between sensitivity and broadness of our view. Hence we need a mosaic to compensate for that.**

Promising complementary approaches include longitudinal and retrospective studies, multimethod assessment of key projects, expert judgment, and documentation.

Longitudinal and Retrospective Studies--In accord with the cumulative and long-term nature of the acculturation process, NSF should experiment with methodologies that seek to understand the long-term impact of informal experiences in the informal setting.

- Longitudinal studies of the developing interests and skills of young people may help shed light not only on the role that informal resources play, but also on the interaction of school and out-of-school experiences.
- Retrospective studies may help uncover common patterns in the development of scientific interests and talent, and they may not only be useful in studying those who become proficient in the study of science, but also help understand how scientific interests are either nourished or discouraged at early ages.

Both longitudinal and retrospective studies can vary in time scale, frequency of measures, and target population (from total population to particular audiences--e.g., those who are exposed to a particular show or exhibit, the science sophisticated, minorities, or women). A mixture of these foci is likely to increase the understanding of the acculturative role that these resources may be able to play.

MultiMethod Assessment of Key Projects--To complement the broad (and low-resolution) view afforded by retrospective and longitudinal studies, several key projects could be assessed much more closely to document and understand the processes of interaction and the impacts of the projects. The recent study of "3-2-1 Contact!" provides an example of this kind of research. It is important that these studies gain as many different perspectives as possible using multiple measures such as those discussed above relative to formative evaluation. It is also important that they adhere to the general guidelines discussed above (e.g., emphasizing the consumer's view, using a variety of sensitive and open-ended measures). In addition, in certain circumstances, more controlled experiments (such as those done in-house by the producers of NSF-supported science television shows) may add yet another perspective to the detailed understanding of an important NSF project in the informal domain.

Expert Judgment--Given the reservations of those most experienced in the domain about the feasibility of accurately and comprehensively assessing the full value of informal science experiences, there may be an important complementary role for expert judgment and criticism. The best instrument may be the implicit wisdom of those who are most familiar with informal science education. As one participant said, "We are in the Stone Age in terms of measurement [in this domain], but we are not Stone Age in terms of expertise." A collection of criticisms from a range of experts, in combination with a statistical understanding of the numbers in the audience that are "reached," may provide a better understanding of the learning opportunity that NSF-funded informal resources are providing than any empirical measure of individual learning outcomes. One participant described the potential of such a combination:

> If there is a program that A. C. Nielsen tells us is reaching X million kids and it is also a program that this community of experts says is a damn good thing to put out there, then I think that the combination of those two facts is powerful.

Documentation--Statistical documentation can provide a picture of who is using informal resources and a rough description of how they are being used. Qualitative documentation can help to provide a more refined picture of the kinds of experiences these resources are engendering. In combination with other types of assessment activities, documentation of key projects can go far to build a meaningful assessment of complex learning activities.

Overall, a Program of Applied Research Is Needed To Search for New Ways To Think About and Approach the Assessment of Informal Learning at the Individual Level

The need now is for a program of applied research that pursues these areas. Progress in assessing and understanding informal learning depends as much on developing and articulating an overarching view of informal science education--its nature, mission, and role--as on actual assessment of NSF-funded activities. More theoretical work to synthesize the work of the field and develop better assessment paradigms is needed. As one participant summarized:

> A small fraction of assessment issues lie in the area where evaluation measures come readily to hand. It seems fairly clear that even in areas where the intent is pragmatic assessment, there is a research level required to design the instrument, to design the means to answer the question. This is a field that clearly does not have a set of standardized measures or instruments. Significant thinking is necessary before you begin....
>
> There is also, I think, still a need for some overarching view of what the nature of the public understanding of science is.... The field itself still has a very primitive notion of its own philosophy. We need more philosophical discourse and a good deal more theoretical thinking. There needs to be more strategy from a societal standpoint where informal science education is positioned. At the level of how informal learning functions, a much wider variety of paradigms is needed. One of the things I've been struck with in this dialogue is how inarticulate we all are, because we don't have very good visual or verbal or other models to describe the transactions and interactions of informal learning. We are imagery poor, we are paradigm poor.

Finally, there need to be more meta-analyses--attempts to synthesize the best of the work that has been done and to combine it with some quantitative and robust information to draw more general conclusions. If you will, we need a strategic assessment based on a selective look at prior research and prior data. This adds up to a substantial recipe for better understanding. I'm not sure it satisfies an internal NSF interest for a report card for how well its programs are doing, but it does rather well in defining some directions and some boundaries for the domain of informal science education.

REFERENCES

Hurd, P. D. (1986). Personal communication.

Zuckerman, H. (1977). *Science elite: Nobel Laureates in the United States.* New York: Free Press.

VI A SYNTHESIS OF LITERATURE: ASSESSING THE INFORMAL SCIENCE LEARNING EXPERIENCE

By

Mark St. John
Patrick M. Shields

VI A SYNTHESIS OF LITERATURE:
ASSESSING THE INFORMAL SCIENCE LEARNING EXPERIENCE

[This paper provided a framework for the discussion at the meeting described in Section V. The paper is written in three sections that correspond to the three major parts of our discussion.]

The first part of this discussion paper describes our understanding of the logic that underlies NSF's initiatives in informal science education. We focus the discussion on individual outcomes hypothesized to result from informal science learning--in other words, what NSF hopes people take away from informal science learning experiences. The first part of the meeting will be spent revising our formulation of the intended (and possible) outcomes of informal science learning.

In the next part of the paper, we describe ways that the outcomes of informal science learning might be assessed. Each link in the chain of logic that underlies NSF's investments can be studied using a range of approaches. We summarize the different approaches that researchers and evaluators have used to date and discuss how their approaches have contributed to an overall understanding of informal science learning. The second part of the meeting will focus on the work done to date and on suggestions of other approaches that look promising.

The final part of the paper discusses a range of factors that must be included in thinking about NSF's assessment plans. In the third and final part of the meeting, we will discuss these considerations and examined how NSF should--and should not--look at the individual's learning experience as a part of its overall future evaluation and research in the informal education domain.

Background

This paper and the meeting for which it was prepared address one aspect of SRI's current work in Phase II of an "Assessment of Initiatives Available to NSF in K-12 Science Education." SRI's project started in 1986, a few years after NSF once again became active in K-12 science education. During the course of the program rebuilding process, Congress required NSF to seek outside assistance in developing a "science education plan and management structure" for the Foundation. As part of NSF's response to that requirement, its Directorate for Science and Engineering Education (SEE) awarded SRI International a contract to perform a large-scale assessment project to be done in two phases:

- The first phase of the project has been completed. The results of that investigative work are described in a set of reports that outline in detail a range of opportunities for NSF's educational investments as well as strategic ways of addressing them (Knapp et al., 1987a, b, c).

- The second phase of the project had the goal of helping NSF find ways to assess its programmatic investments in K-12 science education on an ongoing basis. "Assessment" here is meant in broad terms: it includes descriptive as well as judgmental activities; it includes the study of the contexts in which the investments are made; and it includes the study of all aspects of the investment process--not just the measurement of project outcomes or the achievement of program goals.

The second phase is to produce a report that suggests plans and procedures for assessment. To make these proposed assessment ideas concrete, the second phase includes a pilot test that is experimenting with a range of assessment ideas, focusing on the assessment of investments in informal science education.

The investment strategy of NSF in informal science education can be examined at various levels. On a "macro" level, we are interested in the appropriateness and significance of NSF's role vis-a-vis the full range of informal science activities. At a more operational level, the effectiveness and cost efficiency of the funding strategies that NSF uses to influence the practices of informal science are relevant. Finally, on the level of the individual, we are concerned with the learning outcomes resulting from individuals' interaction with the resources that NSF funds.

The meeting will concentrate on assessment questions on this last level--that of the individual--and this paper reflects that. At the individual level, the issue of assessing "learning" arises. Compared to assessing NSF's national role or the efficiency of its funding strategies, the assessment of informal science "learning" is perhaps the most difficult and the most central issue. Simply put, the question is this:

What do people learn (take away) from their interactions with informal science resources?

There are several reasons that in our pilot test we did not attempt to do a major study of our own to illustrate how this kind of assessment could be done. One is that our resources and time were too limited. The second reason is more profound and humbling--we do not know how to do such a study any better than those who have come before us. From reviewing the limited state of assessment practice in the field, we found a need to rethink the entire issue. Thus, we convened the meeting to "back up a step" and reexamine what approaches do and do not make sense for assessing "learning" in this domain.

The Logic of NSF's Investments in Informal Science Learning

NSF's charter includes the mandate "to strengthen science education at all levels"--a charge that provides the rationale for NSF's investments in the formal educational system as well as its growing investments in the domain of informal science education. Both formal and informal science education investments ultimately are in support of NSF's overall purpose of enhancing the scientific and technological capabilities of the nation.

Through its support for the development and improvement of informal science education resources, the Foundation seeks to improve the general level of "scientific literacy" across the country. To become scientifically literate is in some sense synonymous with becoming more familiar with, and more a part of, the "subculture of science, math and technology." Thus, informal and other investments can be seen as efforts to help acculturate the people of this nation to a world where science, mathematics, and technology are major parts of the overall culture. Paul Hurd, a prominent science educator, put it this way:

> We live in a scientifically and technologically driven economy, and we live in a culture that throughout this century has been identified with science--the atomic age, the space age, the computer age, etc.... The total resources (formal and informal) for science education have become a question of education for living in a science and technology culture. This makes science education a matter of acculturation--so far a little recognized goal.... Formal and informal education are ways (properly conceived) of making it possible for children not to be foreigners in their own culture.... For these reasons I cannot be unbiased about the importance of informal education in the sciences.... (Hurd, 1986)

We believe that the idea of acculturation is very useful in understanding the role that informal science resources are expected to play. Television shows, museums, recreational organizations, newspapers, etc., are prevalent media that provide an ongoing opportunity for people to encounter and learn about the phenomena, ideas, and events of science. Within these media, the specific resources that NSF funds are designed to provide a wide range of rich educational experiences. But what exactly is the outcome of these experiences? What do people "get" from a science television show or a visit to a science museum? And how do these experiences help NSF fulfill its larger mission of enhancing the scientific capability of the nation?

In Figure VI-1 we present our understanding of the logic of NSF's initiatives in informal science education, as it affects individuals. The basic idea represented in Figure VI-1 is that informal science resources provide experiences that lead individuals to feel more knowledgeable about science, more interested in it, and more capable of future learning. In turn, this knowledge, interest, and capacity can alter individuals' future behavior toward science learning. In short, the resources provide experiences that lead individuals to greater acculturation in science and mathematics.

In the sections that follow, we describe in detail each link in the process of acculturation. Because we are focusing on individual learning, we do not discuss the initial link in the investment process in which NSF funding leads to the creation of informal science learning resources.

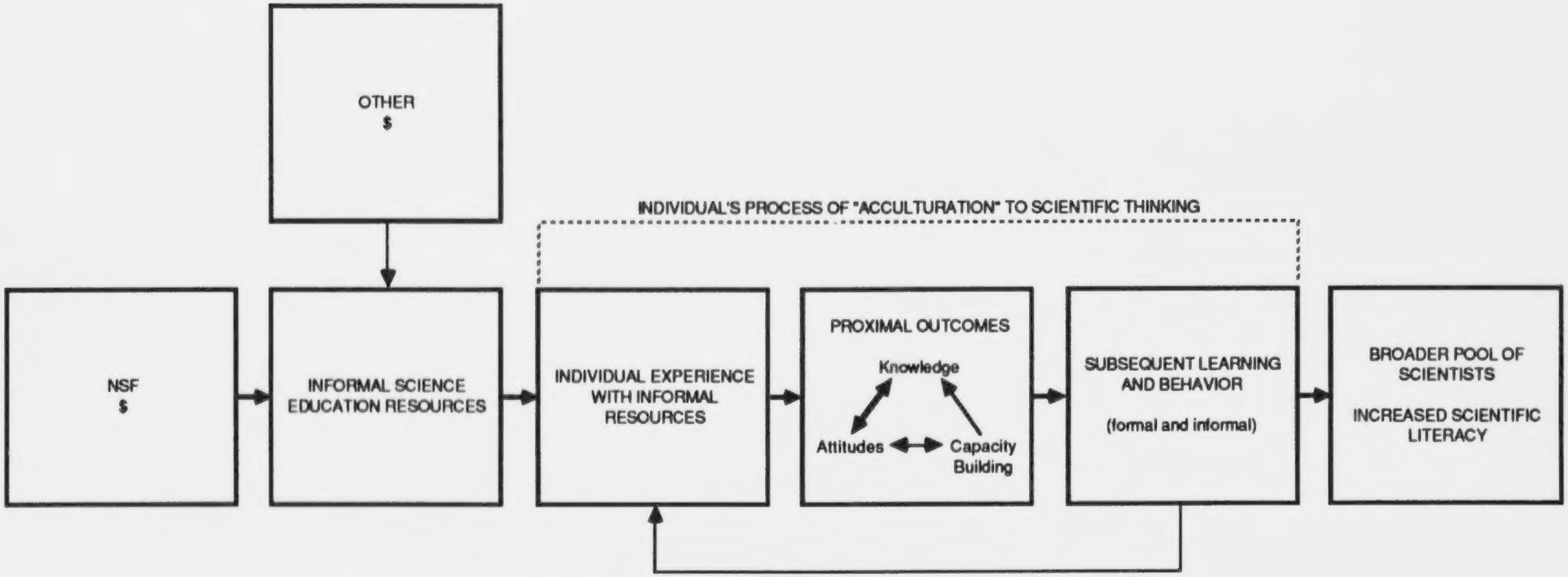

OTHER
$
INDIVIDUAL'S PROCESS OF "ACCULTURATION" TO SCIENTIFIC THINKING
NSF
$
INFORMAL SCIENCE
EDUCATION RESOURCES
INDIVIDUAL EXPERIENCE
WITH INFORMAL
RESOURCES
PROXIMAL OUTCOMES
Knowledge
Attitudes
Capacity
Building
SUBSEQUENT LEARNING
AND BEHAVIOR
(formal and informal)
BROADER POOL OF
SCIENTISTS
INCREASED SCIENTIFIC
LITERACY

FIGURE VI-1 THE LOGIC OF NSF'S INVESTMENTS IN INFORMAL SCIENCE LEARNING (in terms of effects on individuals)

Link 1: Informal Science Education Resources Promote Educationally Valuable Experiences

The first link in the hoped-for chain of events is that NSF's support of informal science education resources, in conjunction with the support of other organizations, will increase the opportunity for a wide range of people to have positive experiences with the ideas and practices of science and mathematics. The rich and diverse range of resources that are funded are intended to promote a corresponding range of experiences for large numbers of people in cost-efficient ways. Also, an important part of the hypothesis is that informal science education resources are thought to provide educational opportunities to a wide array of individuals who are outside of the "scientific pipeline," who might otherwise have very limited access and exposure to "real science."

Link 2: Informal Science Experiences Promote Many Forms of "Learning" About Science

The next link in the chain concerns what we call the "proximal" outcomes of an individual's experience with informal science education resources. (The outcomes are "proximal" in that they are nearer to the origin of the initiative chain, in contrast with the "distal" outcomes of improving the scientific literacy of the country and providing more scientific talent.) In categorizing these outcomes into three broad areas--transmitting knowledge, influencing attitudes, and building capacity for future learning--we have attempted to include the wide range of espoused goals of informal science education. We discuss each in more detail below.

Transmitting Knowledge

Cultural scientific literacy--This term includes the acquisition of basic conceptual and verbal forms of scientific literacy in the sense of being familiar with the important ideas and language of the time. Even a brief introduction to the language of science can empower people to continue to learn more about those ideas, whereas a complete unfamiliarity with the language and terms of science will often prevent any further learning at all. Such scientific literacy is seen as an increasingly important aspect of a greater cultural literacy, as discussed by Hirsch (1987) in his recent essay:

> To be truly literate, citizens must be able to grasp the meaning of any piece of writing addressed to the general reader...the comprehending reader must bring to the text appropriate background information that includes knowledge not only about the topic but also the shared attitudes and conventions that color a piece of writing...much of the required background information is necessarily vague, whether we are conversing or reading. What counts is our ability to grasp the general shape of what we are reading and to tie it to what we already know.... Background knowledge does not take care of itself. Reading and writing are cumulative skills; the more we read (learn), the more necessary knowledge we gain for further reading (learning)....

Civic scientific literacy--Becoming aware of and understanding public policy issues involving the interaction of science and society is another kind of knowledge transmitted through informal science resources. Issues such as the depletion of rain forests, nuclear power, genetic diversity, and ozone depletion are increasingly in the news and are also increasingly the subjects of television programs and museum exhibitions. As well as providing a background to these issues, informal resources are seen as a transmission vehicle for scientists and informal educators to "convey important messages" to the public.

Knowledge of the scientific enterprise--Informal resources can teach about science itself. Scientific biographies, documentaries, and historical exhibitions are all aimed at portraying a picture of the history and the people of science. Beyond the knowledge of the major events and discoveries of science, it is hoped that the public can gain a more accurate picture of the nature of the scientific process.

Fostering Positive Attitudes and Motivation

Excitement about science--Informal science educators place a heavy emphasis on making science accessible, interesting, and exciting for the public. They want people to be engaged with the phenomena, ideas, and processes of science. Some argue that unless learning is overlaid with some level of personal involvement and emotion, it will not be significant or long-lasting.

Motivation for further activity and learning--Provoking curiosity is a prime goal for many informal science educators. Informal resources may be effective not only at generating excitement about a particular topic or phenomenon, but they also may lead to a more general interest in the learning process itself. That is, as a result of successful experiences, people's attitudes about learning or their own use of informal resources may change. They may become more interested and more confident in pursuing a topic on their own using a range of informal (or formal) resources.

Improving attitudes toward science--If informal experiences offer people a positive personal experience, and if they provide a perspective on science as an exciting human endeavor aimed at discovery, then people's attitudes about the scientific enterprise may be improved. In this way, informal science experiences may help to improve the image of science and technology in the public eye by changing people's notions about what science is, how it functions, and what it produces.

Changing the sense of self and science--The "aha" experience (where one is excited about a sudden new understanding) is a sign of success to many informal science educators. Not only does it signify that people have understood something for the first time, but it often indicates that the breakthrough is important and exciting for the person. The discrete nature of these emotional events may be important--the emotional impact of informal resources may not be linear in time but rather it may arise from a single critical event. (Stephen Jay Gould describes the impact of his first encounter with the dinosaurs of the natural history museum in this way.) This idea of "landmark learning"--where a single event in the informal domain can have long-standing impact on someone's relationship to science and even influence career decisions--is complementary to another notion that informal experiences are cumulative and build on each other over time. Whatever the process, one outcome of interaction with informal resources may be the alteration of individuals' ideas about themselves and about their abilities, interests, and capacities for understanding or doing science.

Building Capacity

These outcomes concern developing individuals' capacity to think scientifically, to bring a scientific approach to solving a problem, and ultimately to continue their own learning and development process.

Experiential scientific literacy--In addition to introducing the language, concepts, and abstractions of science, informal resources can help to build a base of direct experience with the objects and phenomenon of science. The kind of learning that takes place occurs in nonverbal modes. In watching images on television and in interacting with phenomena in a museum, a wide range of kinesthetic, haptic, and iconic learning may be taking place. The feel of resonance in pushing a heavy pendulum or the images of a vortex in a tornado exhibit may never be associated with formal scientific terms. These kinds of "primitive" experiences, some argue, provide for a necessary nonverbal literacy--an internal data base of experience upon which more formal conceptual learning can build.

Changes in preconceptions and cognitive structures--People interact with informal resources through the veil of their prior knowledge and existing conceptions of how the world works. Because of the opportunity informal resources offer for direct observation and/or experimentation, they may be effective means for letting people address the inconsistency between what they "know" and what they encounter in the direct observation of phenomena. Learning in this way may offer a chance for people to effect a deep alteration in their existing cognitive structures and mental conceptions of the world.

Acquiring scientific habits of mind--By providing for personal interaction and a chance to "do science," informal resources may help to develop curiosity, confidence, and skepticism--all of which are essential traits of a scientific perspective. Watching a television show in which scientists are shown puzzling over the history of the Mediterranean, or trying to understand in a science museum how a glass-enclosed automobile transmission works provides people with case studies of scientific exploration in which they can become involved. Through experience, they learn about the "scientific method" and may even learn a more generic scientific approach to asking questions and seeking their answers.

Link 3: Informal Learning Experiences Foster Subsequent Learning and Alter Behavior

The experience of interacting with an informal science education resource--watching a science TV show, visiting a science center, participating in a club's activities--cannot be expected in itself to accomplish the broader goal of increasing acculturation or scientific literacy. Rather, the next link in the logic of NSF informal science investments hypothesizes that interaction with informal resources may alter subsequent patterns of learning and behavior. That is, an informal learning experience may change how the individual interacts with other potential science learning experiences--both in and outside of formal educational institutions. The range of changed behavior is, of course, limitless, but would include, for example:

Immediate shifts in behavior--These include actions that result directly from an experience with an informal science resource.

- Turning off the TV and talking with others about the scientific phenomenon discussed on the show.
- Making a connection between something seen in the world with a phenomenon observed during an earlier museum visit.
- Borrowing a book to learn more about a particular phenomenon.
- Committing oneself to a return visit to a science center or to watching the TV show on a regular basis.
- Choosing a topic for the school's science fair.

Intermediate shifts in behavior--These include activities that are somewhat removed from the immediate experience, but that facilitate the individual's scientific acculturation. They are in a sense an institutionalization of short-term changes.

- Regular viewing of a TV show.
- The pursuit of a hobby related to a scientific interest: ham radios, computers, birdwatching, etc.
- Choosing to pursue more advanced science subjects in school.
- Repeated visits to informal educational institutions.

Long-term changes in behavior--These reflect significant shifts in an individual's lifelong interaction with science and mathematics.

- A commitment to pursue a career in a science-oriented field.
- Developing a permanent attentive attitude toward issues of science and technology.
- Employing the habits of mind and a scientific perspective in a wide range of investigative and problem-solving efforts.

The Final Link: Increased Scientific Literacy and a Broadened Pool of Scientists

The final link in the chain represents the aggregation of increasingly acculturated individuals, which leads to an overall rise in the national levels of scientific literacy and a larger pool of individuals aspiring to scientific careers. Obviously, the extent to which the nation's scientific and technological capacity increases depends on numerous resources and factors beyond the informal domain. For the sake of simplicity and because they are beyond the realm of anything NSF can influence, we have not attempted to include these factors in our discussion.

Approaches to the Assessment of Informal Science Learning

Given an understanding of the logic of NSF's investments in informal science learning, we can turn our attention to promising approaches to assessing these investments. The complexity of the logic underscores the impossibility and inappropriateness of exploring direct causal relationships between NSF's investments and the long-term goals of creating a broader pool of scientists and of fostering increased scientific literacy for the nation as a whole. At the same time, explication of the logic that underlies NSF's initiatives highlights key points at which appropriate assessments might be carried out. At each link in the chain, we can look for evidence that short-term learning outcomes have taken place.

In general, the further "downstream" one travels along the chain of logic, the more problematic it is to establish that meaningful relationships exist between NSF-sponsored resources and the relevant outcomes. It is, for example, much easier to judge whether NSF funds resulted in the production of a particular TV show or the extent to which the show attracted an audience of 8- to 12-year-olds than it is to assess the relationship between the show and an individual's subsequent behavior, such as a student's curricular choices in high school. Nevertheless, the importance of a particular relationship might call for difficult and untried attempts at assessment. For example, one might argue that the question of whether informal learning experiences build individuals' capacity for more effective learning in formal settings is so crucial that we must attempt to understand that link. There emerge, then, two primary criteria for choosing to focus an assessment activity on a particular link in the chain: its importance or power in affecting important outcomes and its assessability.

Below, we outline some potential approaches to assessing each link in the chain of logic guiding NSF's informal science learning investments. For each approach, we identify relevant studies in the field, recognizing that many of them simultaneously examine different points in the chain of events. Table VI-1 provides a matrix of the types of study by those links in the chain to which they are relevant.

Approaches to Assessing Link 1 (Informal Learning Resources Encourage Educationally Valuable Experiences)

At this point in the chain of logic, the major questions concern the extent to which individuals take advantage of a particular informal science education resource, the quality and nature of that experience, and the quality and nature of the resource itself. These questions have been addressed by audience (or visitor) studies, qualitative analyses of educational experiences, and expert criticism, respectively.

Audience Studies--On the quantitative side there is a need to understand the informal learning experiences in terms of the numbers of people who interact with the different resources that NSF funds, the demographics of those audiences, and the description of their interaction in quantitative terms (e.g., frequency of interaction, average time interacting). These kinds of questions have been addressed in a wide range of audience (visitor) studies where the researcher focuses on the numbers, composition, behavior, and motivations of those who come to use the resource. In addition, such studies may make rough attempts to gauge the impact of the interaction on the visitor (without going into great detail about the processes of that interaction). See, for example, Brennan (1977); Crane (1987); Dunbar and Borun (1980); Hill (1971); Hood (1983); Loomis (1987); and Miles (1986).

Naturalistic Studies--On the qualitative side there is a need to understand the nature of the experience of people as they use informal science resources. Anthropological and ethological approaches have been used in the museum setting to try to capture in a naturalistic fashion the kinds of interactions that occur between the individuals and the resources as well as between individuals themselves. Examples of such studies include Cone and Kendall (1978); Diamond (1980); Laetsch et al. (1980); and Wolfe and Tymitz (1979).

Connoisseurship and Criticism--The judgments of informal science "experts" may be useful in helping to articulate the design criteria that are important to the success (or failure) of informal resources to foster learning. It is important to acknowledge that this evaluation process is not "objective." It is also important to realize that NSF currently relies heavily on expert critics in its peer review system for evaluating proposals. There is no reason that such criticism could not also be useful in helping to understand, if not measure, the educational value of informal resources. Given the different values and multiple outcomes of informal education, it is unlikely that a process will generate enough consensus to draw conclusions about the overall value of the informal resource being criticized. However, that is not the goal of criticism. Its purpose is to convey to others the educational nature

Table VI-1

APPROACHES TO THE ASSESSMENT OF INFORMAL SCIENCE LEARNING

	Links in the Logic of NSF's Initiatives		
Assessment Approaches	Informal Learning Resources Encourage Educationally Valuable Experiences	Informal Science Experiences Promote Many Forms of Learning	Proximal Outcomes Lead to Subsequent Learning and Behavior Change
Audience studies	X		X
Naturalistic studies	X	X	
Criticism	X	X	
Treatment-effect studies		X	
Process focused studies	X	X	
Longitudinal retrospective studies		X	X

and qualities of an exhibit or television show, with the goal of explaining what the exhibit/television show is, and how it is what it is, rather than to assess what it does. The goal is to illuminate, to educate, and to focus the debate, not to determine overall value.

Such criticism exists in the literature on informal science education in the form of essays. The following are examples of connoisseurs sharing their insights about the general nature of the informal science enterprise: Bodmer et al. (1985); Dackman (1981); Laetsch (1987); Oppenheimer and the staff of the Exploratorium (1986); and Shettel (1973).

Approaches to Assessing Link 2 (Informal Science Experiences Promote Many Forms of Learning)

The central question in the assessment of informal learning experiences concerns what individuals take away from the immediate learning experience. We have already described a series of proximal outcomes, including knowledge gains, attitude changes, and increased capacity for later learning. Over the past 20 years several quite different approaches reflecting the research assumptions of different disciplines have been imported into the informal science arena to assess these kinds of outcomes.

Treatment-Effect Studies--There are many studies that follow a basic research paradigm where informal science resources are seen as the instructional medium and evaluation has the role of determining the efficacy of that instruction. Such studies derive from a treatment-effect metaphor and rely on a basic "input-output" methodology. Following the model of instructional technology, they rely on the stated objectives of the exhibit or television show to focus the assessment of outcomes. Often such studies are done for local purposes with the aim of improving the design of an exhibit or broadcast production. Some examples of such studies are Birney (1987); Borun (1977); Borun and Miller (1980); and Brooks and Vernon (1956).

Basic Research into the Immediate Learning Environment--Other studies aim at measuring proximal outcomes as a part of a more basic research program with the more ambitious aim of discovering generalizable design principles. Such studies hope to generate an empirically based science of informal science education. Examples include Miles and Tout (1978); Miles et al. (1982); Nicol (1969); and Peart (1984).

Process-Focused Studies--The work represented by writings in this area tries to gain a detailed understanding of the processes of informal science learning. Unlike the more experimental "input-output" approaches that seek to measure learning gains as a result of the informal "treatment," the effort here is to describe in detail the nature of the interactions between the visitors--the motivations, scripts, preconceptions they bring to the interaction--and the informal resources.

The works include efforts to build a kind of theoretical understanding of the informal learning experience by examining it with the eyes of an educational psychologist. One set of questions on which this stream of research has focused has to do with cognitive development. See, for example, Duensing (1987); Jensen (1982); and Vukelich (1984).

In addition to a developmental perspective, there are very recent efforts to apply the perspective and tools of cognitive science (expert-novice studies, detailed analysis of protocols) to understand the inner nature of the informal learning experience. The basic focus of such studies is on the information processing of the individuals engaged in using informal resources and the impact of such experiences on their mental representations of the world. In fact, the informal environment may even provide a good research setting for much more basic studies in cognitive science. The following papers discuss this new perspective and its application in understanding the learning of science: Carey (1986); Champagne and Klopfer (1983); and Feher and Rice (1985).

In summary, the assessment of proximal outcomes of informal learning experiences has been attempted from a variety of perspectives with a variety of research approaches. Overall, it is fair to say that the effort to assess these outcomes has been modest and remains relatively unsophisticated. There are few detailed published empirical studies of learning in informal settings. If it is true "that most of what we learn, we learn from informal resources," then the study of such settings is significantly underrepresented in educational research. Also, because the proximal outcomes include abstract and intangible types of learning (e.g., scientific habits of mind, experiential literacy), their assessment has remained undone or even superficial.

Approaches to Assessing Link 3 (Proximal Outcomes Lead to Subsequent Learning and Behavior Change)

This link is largely uninvestigated. Increasingly, a major part of the rationale for informal science investments rests on the assumption that informal science experiences provide an experiential base as well as motivation for further activity and further learning. There are perhaps two approaches that make sense for investigating this assumption:

Longitudinal Studies--Studies of the impact of a range of influences on the development of young people may help to shed light on the role that informal science experiences play in encouraging and supporting scientific activity and interest over the long term. There are few such studies to date, but two examples are Diamond et al. (1987) and Miller (in progress).

Retrospective Studies--Here the notion is to work from the other end of the hypothesis chain--that is, to start with individuals' present degree of literacy or acculturation and trace "backwards," searching for the influences that encouraged their acculturation. The role of informal science experiences can be investigated as a part of this broader search, both as to their overall impact and as to the way in which they interact with other influences (e.g., parents, teachers, formal education). Not all studies need focus on the scientifically accomplished. Some of these studies might focus on the general population and examine the sources of their learning (e.g., Miller, 1987a, b; Riccobono, 1986). Others look solely at those who have developed a high level of interest and involvement in mathematics and science (e.g., Bloom, 1985; Zuckerman, 1980). In a very small pilot study at SRI, we have been experimenting with this retrospective approach, interviewing a range of people involved in the science and mathematics enterprise (see Section VII).

The important aspect of both the longitudinal and the retrospective approach is that they adopt a much wider view--asking about the cumulative long-term influence of informal science experiences as opposed to the outcomes of a single viewing of a television show or museum visit. In this way, the approaches seek to validate the hypothesis of the initiative, not by aggregating the effect of individual interactions with informal resources, but rather by looking at the net cumulative effect over many years.

Toward an Agenda for Assessing Informal Science Learning at the Individual Level

So far we have laid out a detailed description of the rationale for NSF's investments in terms of individual learning and of the ways in which this learning might be assessed. Up to this point, we have presented a large-scale and undifferentiated map of the assessment terrain. The map of assessment outcomes and approaches should be of some use to NSF in understanding the issue of assessing individual learning, but ultimately we would like to highlight certain sections of this map, pointing out areas that should be pursued and areas to stay away from. In this section of the paper, we try to set a frame for a discussion that will identify the most promising assessment opportunities for NSF as well as the most (and least) promising assessment approaches to address them.

We do not intend to develop a detailed assessment plan for NSF. Rather, we wish to see where assessment efforts on the individual level can best serve NSF, informal science educators, and others. Conversely, we want to see if there are approaches that are likely to be unproductive or even inappropriate. Our goal is to form an outline of an assessment agenda by identifying promising areas and approaches for:

- Evaluating the contributions of NSF projects, initiatives, or programs to informal science learning.
- Sponsoring basic research efforts and studies that focus on the individual learning from informal science experiences.
- Conducting meta-level activities (like the meeting for which this paper was prepared) that continue to explore the issue of how best to assess informal learning.

Factors Affecting the Assessment Agenda

A number of factors, many outside the control of NSF staff, will influence particular choices in the Foundation's research agenda. These include the constituencies for which the assessment is intended, the media in which the assessment takes place, and practical constraints facing NSF staff. We discuss each briefly below.

Constituencies--Assessment does not happen in a vacuum, and the needs and perspectives of the different constituents should help to shape assessment efforts. Different constituencies have different interests in and uses for the information that is generated by any assessment effort within the Foundation. Moreover, they often want assessment done for different purposes. Different constituents also value different criteria when they examine the credibility or value of assessment information. The main constituents to be served by assessment information on the individual level are:

- *NSF policymakers at all levels.* The NSF staff are the central audience for any assessment activities. Assessment results provide important information for both reporting/justification and planning purposes. As the Foundation increases its resources in the informal domain, it needs to examine the empirical evidence available to answer the "micro-level" question: What do individuals get from the resources that NSF funds? In addition to justification, the specific understandings gained from studies of informal learning may help to inform the future initiatives in the informal domain. For example, the recently completed study of "3-2-1 Contact!" (Crane, 1987) raises some interesting questions about the age distribution of the audience, what these different ages are learning, and how their numbers are changing. Such information can help form future policy.

- *Informal science educators.* Those who are involved in carrying out NSF projects in informal science education may also benefit from improved understandings of what people are learning from the resources they design. There are different degrees of sophistication with which practitioners study learning in their own environments. The Children's Television Workshop, in fact, has institutionalized a process of studying children's responses to their shows as a means of on-line checking and reshaping their children's television productions. Museums, to a lesser extent, systematically use formative feedback to improve their exhibit designs.

- *Educational researchers.* Educational researchers and cognitive scientists may find the setting of informal institutions to be very good for pursuing studies of cognitive development. Because informal learning is intrinsically driven (as opposed to extrinsically driven in the formal setting), the informal setting may provide a valuable arena for conducting basic educational research. Basic research studies that focus on individuals in informal settings may be very useful in understanding what attracts people to learning about science and technology, what stimulates their curiosity, and what helps them to continue learning.

- *Congress and the public.* The state of the nation's scientific and technological health is drawing increasingly concerned questions from Congress and the public. Similarly, the K-12 educational system is under scrutiny and is the subject of much debate. These larger national concerns provide a context in which information about learning in informal settings is of increased importance. Not only is there a need for empirical information, but also there is a need for a clear articulation of the role(s) that informal learning plays vis-a-vis these larger issues.

Media--The different media offer different comparative advantages and limitations for assessment activities. Television is a visual medium, ideal for linear storytelling and for providing a window on parts of the world that are otherwise inaccessible. Museums offer actuality, with the chance to interact with a wide range of interesting objects and phenomena. Print is a highly flexible and inexpensive medium, effective in the verbal and visual domains. These differences should be part of the thinking that helps to shape assessment opportunities and approaches.

Practical Constraints--NSF faces several constraints in thinking about its investments in the assessment of informal science learning. There are financial constraints--assessment efforts will always represent a small part of the overall investment in informal science. Also, there are political constraints--different constituencies have politically significant relationships with each other and are wary of the use of assessment information. Finally, there is limited expertise available in the field to carry out whatever assessment plans the Foundation may decide upon.

These considerations--the constituencies involved, the nature of the media, and the practical constraints--will influence efforts to outline a practical assessment agenda for the Foundation. They will help shape our attempts to determine when and where it is appropriate and productive for NSF to include the assessment of informal science learning in its overall assessment plans.

Illustrative Guiding Principles

Finally, to provide an example of the kinds of guiding principles we might arrive at as a group, we offer the following postulates about the assessment of informal learning:

- There are multiple and very different modes of learning that may be occurring in informal educational environments. Learning must be very broadly defined if assessment is not to miss much that is essential to the informal learning experience.

- Many of the most important learning outcomes are hard enough to define--let alone measure. What is easy to measure is often trivial; what is most important is often very difficult to assess.

- Strict adherence to research and evaluation approaches used in the formal education domain is likely to be inappropriate in the informal education domain.

- The informal domain is diverse in every dimension--the participants, their motivations, their backgrounds, and the kinds of resources they encounter. The goals and styles of the informal educators are diverse. All of this implies the need for multiple ways of assessing what is "learned."

- The assessment of informal learning cannot be solely goal driven. The articulation of goals may help determine some areas to focus on in an assessment, but they should not be the sole determinant, as the intent of the resource may play a small part in what the intrinsically driven participant takes away.

- There is a kind of uncertainty principle operating in the informal domain--the observation affects the observed. Methods of assessment that impose conditions that are not part of the informal environment may badly distort that which is being assessed.

- There are multiple functions that assessment of learning may fulfill and multiple audiences who care. There is no single kind of assessment that will satisfy all functions and all audiences.

At a more specific level, we hope the meeting will generate suggestions about particular links in the chain of events from NSF funding to ultimate outcome that could be productively addressed by specified approaches. These suggestions could specify productive variables, settings for assessment, ways to orient findings to particular audiences or media, etc.

REFERENCES

Birney, B. A. (1987). *The "Flying Walk" exhibit: Formative evaluation aids in the development of an interactive exhibit in an informal learning setting.* Chicago: Chicago Zoological Society.

Bloom, B. S. (Ed.). (1985). *Developing talent in young people.* New York: Ballantine Books.

Bodmer, W. F., et al. (1985). *The public understanding of science.* London, England: Royal Society.

Borun, M. (1977). *Measuring the immeasurable: A pilot study of museum effectiveness.* Washington, DC: Association of Science-Technology Centers.

Borun, M., & Miller, M. (1980). *What's in a name?* Washington, DC: Association of Science-Technology Centers.

Brennan, T. (1977). *Typical zoo visitor social group behavior.* AAZPA National Conference, San Diego, California (September 18-22). (Available from AAZPA - American Association of Zoological Parks and Aquariums, Oglebay Park, Wheeling, WV 26003.)

Brooks, J.A.M., & Vernon, P. E. (1956). A study of children's interests and comprehension at a science museum. *British Journal of Psychology, 47,* 175-182.

Carey, S. (1986). Cognitive science and science education. *American Psychologist, 41*(10), 1125.

Champagne, A. B., & Klopfer, L. E. (1983). Naive knowledge and science learning. Paper presented at the annual meeting of the American Association of Physics Teachers, New York (January).

Cone, C. A., & Kendall, K. (1978). Space, time and family interaction: Visitor behavior at the Science Museum of Minnesota. *Curator, 21*(3), 245-258.

Crane, V. (1987). *An exploratory study of "3-2-1 Contact!" viewership.* Chestnut Hill, MA: Research Communications, Ltd.

Dackman, L. (Ed.). (1981). Science media conference. *The Exploratorium, 5*(2,3).

Diamond, J. (1980). Ethology in museums: Understanding the learning process. *Roundtable Reports, 7*(4).

Diamond, J., St. John, M., Cleary, B., & Librero, D. (1987). The Exploratorium's Explainers Program: Long-term impact on teenagers of teaching science to the public. *Science Education, 71*(5), 643-656.

Duensing, S. (1987). Science centres and exploratories: A look at active participation, in *Communicating Science to the Public.* New York: John Wiley & Sons.

Dunbar, N., & Borun, M. (1980). *The science museum audience.* Washington, DC: The Association of Science-Technology Centers.

Feher, E., & Rice, K. (1985). Development of scientific concepts through the use of interactive exhibits in a museum. *Curator, 28*(1), 35-46.

Hill, C. A. (1971). An analysis of the zoo visitor. *International Zoo Yearbook, 11*, 158-165.

Hirsch, E. D. (1987). *Cultural literacy.* Boston: Houghton Mifflin.

Hood, M. G. (1983). Staying away: Why people choose not to visit museums. *Museum News, 61*(4), 50-57.

Hurd, P. D. (1986). Personal communication.

Jensen, N. (Ed.). (1982). Children, teenagers and adults in museums: A developmental perspective. *Museum News, 60*, 25-30.

Knapp, M. S., Stearns, M. S., St. John, M., & Zucker, A. (1987a). *Opportunities for investment in K-12 science education: Options for the National Science Foundation--Summary report.* Menlo Park, CA: SRI International.

Knapp, M. S., St. John, M., Zucker, A. A., Needels, M., & Stearns, M. S. (1987b). *Opportunities for investment in K-12 science education: Options for the National Science Foundation. Volume 1: Problems and opportunities.* Menlo Park, CA: SRI International.

Knapp, M. S., Stearns, M. S., St. John, M., & Zucker, A. A. (1987c). *Opportunities for investment in K-12 science education: Options for the National Science Foundation. Volume 2: Groundwork for strategic investment.* Menlo Park, CA: SRI International.

Laetsch, W. M. (1987). A basis for better public understanding of science. In D. Evered & M. O'Connor (Eds.), *Communicating Science to the Public (Ciba Foundation Conference).* New York: John Wiley & Sons.

Laetsch, W. M., Diamond, J., Gottfried, J. L., & Rosenfeld, S. (1980). Children and family groups in science centers. *Science and Children, 17*(6), 14-17.

Loomis, R. J. (1987). *Visitor evaluation.* Nashville, TN: American Association for State and Local History.

Miles, R. S. (1986). Museum audiences. *International Journal of Museum Management and Curatorship, 5,* 73-80.

Miles, R. S., Alt, M. B., Gosling, D. C., Lewis, B. N. & Tout, A. F. (1982). *The design of educational exhibits.* London, England: Allen and Unwin Publishers.

Miles, R. S., & Tout, A. F. (1978). Outline of a technology for effective science exhibits. *Special Papers in Paleontology, 22,* 209-224.

Miller, J. (In progress). *A longitudinal study of the development of adolescent and young adult attitudes towards and knowledge about science and technology.*

Miller, J. D. (1987a). *The roots of scientific literacy.* Paper presented to the St. Louis Sigma Xi Symposium on Science Education (October 6).

Miller, J. D. (1987b). Scientific literacy in the United States, in *Communicating science to the public.* New York: John Wiley & Sons.

Nicol, E. H. (1969). *The development of validated museum exhibits* (Final Report). Boston: Children's Museum.

Oppenheimer, F., and the Staff of the Exploratorium. (1986). *Working prototypes: Exhibit design at the Exploratorium.* San Francisco: The Exploratorium.

Peart, R. (1984). Impact of exhibit type on knowledge gain, attitudes and behavior. *Curator, 27*(3), 220-237.

Riccobono, J. A. (1986). *Out-of-school learning among children, adolescents, and adults.* Report of findings from the 1985 Home Information Technology Study (HITS), prepared for the Corporation for Public Broadcasting and Center for Statistics.

Shettel, H. (1973). Exhibits: Art form or educational medium? *Museum News, 52*(1), 32-41.

Vukelich, R. (1984, Fall). Time language for interpreting history collections to children. *Museum Studies Journal,* pp. 43-50.

Wolfe, R., & Tymitz, B. (1979). *Do giraffes ever sit? A study of visitors' perceptions at the National Zoological Park.* Washington, DC: The Smithsonian Institution.

Zuckerman, S. (1980). The rise of zoos and zoological societies. In S. Zuckerman (Ed.), *Great zoos of the world,* pp. 3-26. London, England: Weidenfeld & Nicholson.

VII RETROSPECTIVE ASSESSMENT DESIGNS: A FEASIBILITY STUDY FOR INVESTIGATING THE LONG-TERM IMPACT OF INFORMAL SCIENCE LEARNING EXPERIENCES

By

Patrick M. Shields
Debra M. Shaver
Mark St. John

VII RETROSPECTIVE ASSESSMENT DESIGNS: A FEASIBILITY STUDY FOR INVESTIGATING THE LONG-TERM IMPACT OF INFORMAL SCIENCE LEARNING EXPERIENCES

The purpose of this pilot activity was to explore the potential of a retrospective approach for studying the long-term impact of informal science learning experiences. In particular, we focused on seeing whether a retrospective approach could tell us anything about the role informal science resources play in contributing to individuals' interests in science and, more specifically, in influencing them to pursue careers in science- and technology-related occupations.

Highlights

It is important to emphasize that the scope of our pilot study and the size of our sample (20 individuals) were so small that we are unable to draw any defensible conclusions about any population, or about the role that informal science resources play in the development of scientific skills and interests. However, many of the individual interviews in our study showed patterns in science career development that are compatible with existing work in this area. These patterns illustrate what might emerge from more extensive and rigorous assessments using a retrospective approach. For example, we saw evidence of:

- Parents and families exercising great influence over career decisions.
- A central role model (often a parent) providing active encouragement for academic achievement and playing an important part in forming respondents' pursuits of science.
- Respondents pursuing studies in science and mathematics as an extension of long histories of successful school achievement.
- A number of the respondents--especially the women and minorities--facing substantial barriers in pursuing their desired careers.
- A strikingly wide range of informal science learning experiences contributing to interests in scientific careers. Such experiences appeared less to be responsible for initiating such interests than to be important factors in sustaining interests already developed.
- Experiences with the type of informal science resources funded by the National Science Foundation (e.g., museums, television) being mentioned by about half of the participants as contributors to their science career development.

- Few differences between minority and nonminority respondents or between men and women in the types of factors that most strongly affected their career choices.

The Pilot Study Approach

This pilot activity tested the feasibility of adopting a retrospective approach for examining the long-term, residual influences of programmatic investments. Such assessment studies begin with an individual (or phenomenon) who might have been influenced by science education resources of the type funded by NSF and seek to identify retrospectively the sources of influence on the individual (or phenomenon). In essence, then, this form of research assesses investments in reverse order, by starting with long-term outcomes and tracing backward through the hypothesized chain of events leading to them.

This approach has been used by others to study the development of scientific interests and talent--for example, in studies of the influences on mathematicians and scientists (e.g., Gustin, 1985) and in research on the development of Nobel Prize winners (Zuckerman, 1977). Retrospective approaches have been used in other fields as well. Such a strategy was used in a recent assessment of NSF's investment in engineering-related research (NSF, 1987). That study began with the most significant advances in the field of engineering over the past 20 years and then traced back to uncover which organizations had funded the research leading to those advances. A variant of the study has been used to identify policy influences on the local implementation of education reform programs and to suggest, through "backward mapping," the kinds of policy action that would facilitate local reforms (Elmore, 1980).

In our pilot activity, we pursued a similar retrospective strategy, but focused on individual scientists' career choices as the relevant outcome. Since we were working from the scientists' current occupational position and were investigating the nature of their history vis-a-vis science, we had to adopt an open-ended approach to include all types and sources of relevant experiences--whatever factors had significantly influenced them. Thus, we started our search with a broad lens, and within what we found, we looked more carefully for evidence specifically about the impact of informal science resources. (The interview protocol appears in Exhibit VII-1 at the end of this section.)

We chose as our subjects a group of technicians and scientists. By asking each to recount their process of developing an interest in a science-related occupation, we sought to answer a number of specific questions:

(1) What influences appeared to be most important in helping to shape these individuals' career choices?

(2) To what extent did out-of-school experiences with informal science education resources influence these individuals?

(3) Is it possible to learn anything about the nature or relative size of the influence played by museums, television, and other informal science education resources on individual career path choices?

(4) Are there discernible differences between men and women or between members of different ethnic groups in terms of events affecting their career decisions?

Rationale for a Retrospective Approach

An informal base of retrospective data has existed for years in the stories that scientists tell about the experiences that helped form their interest in pursuing mathematics and science. Physicist Richard Feynman eloquently describes how in fixing radios as a child he learned to love problems that were "puzzles"--that required thinking, experimentation, and persistence (Feynman, 1985). Other physicists mention the role that institutions such as the Museum of Science and Industry in Chicago played in their youth; a prominent scholar of evolution refers to the momentous occasion of meeting his first dinosaur in a museum of natural history; a biologist reports that he and a number of his associates grew up on farms where they developed a lifelong interest in nature; many scientists speak fondly of the electronics they learned from their model trains, and so on. There is then a substantial folklore about the important contribution that a wide variety of experiences with the world can make to the creation of scientific ability and interest. This folklore adds to the conventional wisdom that informal science education resources play an important, if hard to define, role in the overall learning and understanding of science by the nation's citizens. In our retrospective study we were interested in seeing to what extent we could more systematically explore and document the kinds of experiences represented by these anecdotes.

In Sections V and VI of this volume, we pointed out that the goal of helping people develop a lifelong interest in ideas and issues that are scientific or mathematical in nature requires a long-term process of acculturation. When one begins to live in a new culture, it takes years of living--and a variety of informal experiences--before one feels knowledgeable about and part of that culture. If it is indeed true that becoming scientifically literate is somehow analogous to a long-term acculturation process, then we can see how informal science education experiences might help, in a cumulative fashion, to develop skills, attitudes, and interests in science. Rather than asking what someone learned from an exhibit or a television show, the most important question for assessment then becomes, "What experiences were important in the process of scientific acculturation?" The shift is from a short-term to a long-term view, and from a linear cause-and-effect model of learning to one that is nonlinear, complex, and highly interactive. We believe that a retrospective approach (similar to that used by an oral historian) may provide a way to explore the answer to that question.

There are other reasons to consider experimenting with this rather different and nonexperimental retrospective approach. One goal of nearly all NSF's investments in science education is to broaden the pool of future scientists and technicians. Traditional assessment procedures begin with an NSF-funded resource or activity and attempt to trace its immediate effects, but they rarely can address the issue of long-term impact. The complexity of NSF's investment hypotheses and the myriad social forces affecting individuals' decisions, however, make it nearly impossible to demonstrate how the Foundation's actions have affected broad societal phenomena, such as the numbers of scientists and engineers.

For example, a recent study of "3-2-1 Contact!" (Crane, 1987) found that large numbers of children watch the TV show, that many learn from it, and that some carry out related activities immediately following the show (e.g., initiating discussions or attempting experiments). These are positive findings, in part because we expect that children who watch the show and learn from it will be more likely to pursue other science-like activities in the future and that a greater percentage of children will continue on in science-related career paths. But the Crane study cannot address either of these longer-term outcomes. A longitudinal study of the children that traced their schooling and career choices over the next decade might provide such information, but it would be immensely costly and would yield results only after many years.

Like longitudinal studies, retrospective studies adopt a long-term view. Studies that rely on retrospective recall are admittedly less precise and accurate than longitudinal studies in documenting the evolution of changes, but they do explore the important residues of experience within an individual. More generally, a retrospective approach adds another perspective and tool to the assessment repertoire; this tool yields a different sort of information for the "mosaic" of evidence that is so important in assessing informal science learning (see discussion in Section V).

Pilot Study Methods

To test the feasibility of this research strategy, we wanted to study a sample of individuals who had all become professionals in scientific and technical fields and yet who varied in the circumstances of their upbringing--education, economic, and cultural backgrounds--as well as any other factors likely to influence the choice of a science career. For this reason, we selected a nonrandom sample of informants who were engaged in a range of scientific or technical occupations and who varied in age, gender, and ethnicity. The sample consisted of 20 staff scientists and engineers employed in California-based technological and scientific institutions in the San Francisco area. Computer science and electrical engineering were the two most common fields. Biology and physics were also represented by a number of respondents. The demographic characteristics of the sample are as follows:

Gender		Age Range		Ethnicity	
Male	11	24-29	7	White	13
Female	9	30-35	6	Black	3
		35-44	5	Hispanic	4
		45-55	2		

We considered augmenting the sample to include one further characteristic, essential for a more extensive and complete examination of this topic: the choice of nonscientific occupations. We dropped this important group from the sample because of the limited time for conducting the study and because it was unlikely to lead to substantial changes in our interview research approach. Nonetheless, a complete feasibility study, to say nothing of a full-scale investigation, would have included such a group in the sample.

Open-ended interviews were conducted with each individual for one-half to one hour at the individual's place of business. Interviews were structured to trace the individual's interaction with science and technology back through time (see Exhibit VII-1 at the end of this section). Thus, the first questions dealt with the individual's present occupation, hobbies, and interests. The interviewer then asked questions about the individual's postsecondary education, then secondary school, elementary school, and finally about early childhood experiences. At each point in the questioning, theinterviewer probed for sources of influence on the individual's interests and choices. Why had the respondents chosen to work as an engineer or scientist? Why had they chosen a science major in college? What had influenced them to study mathematics and science in high school? The interviewer encouraged individuals to recount their stories at length, but always pushed them to explain the "why" behind their choices and interests.

After each interview, summaries were written detailing the present position of the interviewees, their educational and personal history, and the factors they identified as being most influential in their career choices. At both the midpoint and the end of the data collection period, analytic summaries of the findings were written that identified themes common to the individual interviews and addressed our specific research questions. These summaries form the basis for the following section.

Patterns of Influence

Once again, we reiterate a disclaimer that our "findings" do not prove anything in a statistically valid way. Our small, nonrandom sample does not allow us to generalize the results of this study to any specified population, nor was this ever our intention. Rather, the aim for this study was to test the feasibility and utility of the retrospective approach for learning about the informal science domain. The patterns that we could see in 20 respondents, we believe, nevertheless raise some issues that are interesting, thought provoking, and deserving of further exploration. We describe below the themes that we saw running through our interviews.

Individuals report that parents and families had a strong influence on career decisions. Consistent with similar research (e.g., Gustin, 1985), the scientists we interviewed, to a large extent, grew up in family environments in which parents were highly motivated and academic achievement was valued. The scientists' parents strongly supported and encouraged their learning efforts both in and out of school. In nearly every case, the scientists reported a childhood and adolescence in which parents held extremely high expectations and in which there was never any question about whether they would go on to college. As one respondent noted:

> My parents had gone to college, as had my older brothers and sisters--there was never any question that I would follow in their footsteps. Attending college and pursuing a successful career were "givens" in my home.

Parents established a scenario, which the child more or less internalized, about the academic and professional future of the child. Interestingly, the scenario was often not science-specific but rather more general. As one respondent said:

> College was never a question; it was always assumed. But there was no pressure to do math and science. The direction I wanted to go in was pretty much up to me.

In his research on the development of research mathematicians, Gustin summarizes the nature of the home environment in a way that is compatible with what we heard from the scientists we talked to:

> The wide range of home environments notwithstanding, intellectual and academic achievement were valued highly in the home.... Models of cognitive and intellectual behavior were available.... The parents of the mathematicians did not intend to impose their own interests on their children.... There was no effort on the part of the parents to emphasize mathematics, neither was there any effort to avoid it...[but] they played a very significant role in determining the manner in which those interests were pursued.... And the first signs of curiosity--questioning, wondering, wanting to know--were encouraged and nurtured. (Gustin, 1985)

Not surprisingly, the scientists in our survey generally came from middle- and upper-class backgrounds. All but three came from families in which one or both parents held professional positions.

By providing role models and specific encouragement, parents (and sometimes others) played an important part in allowing respondents to develop identities as scientists. Parents often served as influential role models for these future scientists. A majority of our respondents had parents (usually the father) engaged in science-related occupations. As children, then, these respondents did not view scientific occupations as "weird" or "too difficult"; rather, they saw them as "what dad and his friends do." One interviewee noted that all of his parents' friends were engineers and that as a young child he had had the impression that engineering and science were things most grown-ups did.

In other cases, parents provided more direct assistance to their children. One respondent noted that her father brought an early version of a personal computer home from the office each weekend. She and her friends would play with it all day Saturday and Sunday. When she enrolled in an all-male computer science course in high school, she was far ahead of her peers and evidenced little fear of machines or mathematics. Another scientist pointed out that his father, who was an engineer at a local aerospace company, had helped him to get summer jobs there. That experience enabled him to get into a prestigious engineering program, which ultimately led to an engineering position with a large computer firm.

Even those parents who could not lend such direct assistance, however, often played key roles in supporting their children's interest in science. One Hispanic women who is now an engineer retold stories of her childhood in which her father, a Mexican immigrant and blue-collar worker, would create mathematical games for her and her brothers to play. The whole family would spend hours together trying to solve puzzles created by her father. In sixth grade, this same woman had the opportunity to interact with a local university professor who came to her school. The professor encouraged the young girl to pursue her dreams of going to college. Sensing the motivating effect of the college professor, the father asked the professor to come to their home once a week to work with his daughter--for which he paid a small sum.

In the few cases in which parents could not provide the needed support or role model, others stepped in to fill that gap. One scientist's father was a truck driver who supported his child's efforts but did not provide for science-related learning resources and opportunities. The child was fortunate enough, however, to have a next-door neighbor who worked as an astronomer at the local university. The astronomer took the child under his wing, brought him on trips to the observatory, taught him how to observe stars through a telescope, and helped to ignite an interest in science that stayed with the child for the rest of his life. In other cases, teachers played similar roles for these future scientists.

In addition to role models, most interviewees could recall specific instances of direct personal encouragement and recognition for their scientific or mathematical ability. At least as important as academic success was the suggestion of an admired teacher that the young student was good in science and mathematics and should pursue it. Respondents told stories of how they had regularly won praise from teachers and administrators for their superior performance. Such positive feedback (even one critical incident of it) prompted many of the scientists to think most seriously about science-related careers, where they thought they could continue to do well. One scientist remembers being told that "with his talents, he should of course have a career in math or science." Similarly, a few of the scientists we talked to remembered a very discouraging time when they were told that they "probably wouldn't want to go into math."

Respondents were encouraged to pursue science and mathematics by long histories of successful school achievement and the promise of economic success in the field. With a single exception, the scientists we interviewed had excelled in school since early childhood. The great majority had enjoyed that success in the mathematics and science subjects. A number of different kinds of rewards and successes helped the young scientists to validate an image of themselves as scientists. Many were placed in Advanced Placement classes where they were given extra support and encouragement to continue on in science. One interviewee told us that "my very high scores on the Math SAT test sealed my destiny." Another spoke of not being able to turn down a chemistry award and scholarship. A woman who returned to school after many years did very well in an introductory computer programming course, much to her surprise, and continued her study largely on the basis that she could succeed at it.

In fact, three-quarters of those we interviewed remained on a scientific "fast track" throughout school and their subsequent careers. These scientists focused on science and mathematics in high school, performed extremely well on standardized tests in these subjects, chose colleges based on their science or mathematics programs, majored in their chosen discipline, and went directly into graduate school or a job with a highly rated company. Some told us that the chance for a good job was a major underlying factor in choosing and persisting on the career track they had followed. Only a handful of our respondents had seriously flirted with majoring in other areas or pursuing careers in nonscientific fields.

Nevertheless, a number of our respondents--especially the women and minorities--faced barriers in pursuing their desired careers. The success and direct career paths of these scientists obscure some of the formidable barriers they had to overcome. A number of the women noted that they had been seriously discouraged at different points in their careers. One said, "At the beginning of high school, a female teacher took me aside and counseled me to steer away from the difficult math and science courses the boys took." In that case, however, she drew support from a guidance counselor she had met who encouraged her to stick with her plans. In other cases, schools made it difficult for students to pursue more than one interest, as one woman noted:

> The Advanced Placement courses in my school were scheduled so that sciences met at the same time as humanities and social sciences. Girls tended to go to the "soft" subjects like English, boys to the "hard" sciences. Right then I knew that I was going to have to give up some of my interests if I were to pursue a career in science. I chose the AP calculus class over English.

In another case, a black computer scientist noted that when he moved from elementary to junior high school and then from junior high to high school, he was regularly demoted from the honors mathematics and science classes and placed in a lower track.

Each time his parents had to come to school and argue with the principal to get him reinstated in the honors program.

Fortunately, a number of the women and minorities in our sample were able to take advantage of special programs designed to facilitate their entrance into science-related careers. For example, one woman noted that throughout high school she had received little encouragement to continue her dream of becoming an engineer. Then, in her senior year, she learned of a pre-engineering course for girls held on the weekends at a local engineering firm. The teachers in the program were very supportive and, importantly for her, she was able to see the connection between the mathematics and science she was learning in school and "real" engineering. After that experience, she noted, "there was no holding me back." In another case, a black male recounted the difficulties he faced as one of a small percentage of minorities at a prestigious engineering university. After his first semester he was faltering in the unfamiliar atmosphere and feared he would not make it. He then joined a special support group of minority engineering students, most of whom had faced difficulties similar to his. With the help of his support group, he was able to pull through the first year and continue on to graduation and get a good job with a well-respected engineering firm.

A wide range of informal science learning experiences played a key supportive role in shaping scientific career interests, but did not appear to initiate or be the driving force behind the development of these interests. All of our respondents enjoyed science-like activities outside of formal schooling as children. Common experiences included playing with a chemistry set, engaging in mathematics games with parents, helping a parent fix an appliance, or reading a science-related book. In fact, what is striking about the interests these people had as children is the range of activities they were engaged in. Listed below are activities that were mentioned in our 20 interviews:

model planes
cooking
ham radio
music
photography
computers
antiques
ballet
coin collecting
chemistry
symphony books
carpentry
electronics
travel
model rockets
aviation
physics/philosophy
stamp collecting
boy scouts
models
rifles
debate
chess

It is clear from our interviews that both parents and children had many different kinds of interests and were engaged in many different kinds of activities.

We also found some differences between men and women in terms of their informal experiences as children. Almost all the men in the sample reported extensive "tinkering" with a variety of toys ranging from building blocks to chemistry sets as young children. In contrast, few women reported such experiences as children. Women's early interest in science tended to be more abstract, less hands-on, and more oriented toward book learning.

We can speculate in several ways about the meaning of these findings about informal "science" activities:

(1) The activities were important not so much because they taught basic science but because they fostered positive attitudes toward science-like pursuits and they provided experience in carrying out personal inquiry and activity. Thus, what these children were interested in may not have been so crucial as the fact that they were interested at all. The value of being intellectually interested in things was one that was clearly part of the family environment and manifested itself in many hobbies and activities.

(2) In each interview, scientists were able to look back on their childhood and find evidence of their science interest in out-of-school activities. It seems that these experiences were much less important for the concrete knowledge respondents picked up than for the legitimacy they lent to science-like activities. That is, for the scientists we interviewed, their informal experiences as children helped them to build a respect for science as an appropriate activity for them. It was important that science was accepted as an "okay" thing to be doing. This support as young children helped respondents to overcome some of the negative stereotypes of scientists as they entered junior high school and high school.

(3) For most young people, it may be that informal science resources do not serve as "entry points" to the culture of science and scientists. It may be that supportive interaction with key individuals (parents, teachers, mentors) and success in school are the real "gatekeepers." The role of informal resources may be most important in providing nourishment for an appetite that is already whetted. For some scientists, however, it is clear that from early on they have developed a style of learning that is highly independent, active, and experimental. As one scientist told us: "I have always had a 'how does it work' mentality. At a very early age I took apart the kitchen clock...and I think that the best teacher is one that leaves me alone to learn on my own." For such individuals, informal science education resources may play a very important role.

(4) One implication of our finding here is that it may not be as important that young students be doing "junior science" activities (e.g., little lab experiments) as it is that they develop an ability and propensity for a wide range of activities. The implication for NSF would be that it may be important to take a broader view of the kinds of "science" activities that ultimately develop interest and talent in the field.

Approximately half of the respondents mentioned experiences with the type of informal science resources funded by the National Science Foundation. Reading was probably the most important activity. The *National Geographic* was mentioned by many participants. Also, some said that they had learned to enjoy reading "technical" things (e.g., magazines like *Popular Science*) and books about physics and philosophy.

Other media played a less important role for most scientists. Those who mentioned television said that their family had valued public television, and that science and nature had come as a part of an interest in television devoted to educational topics. "Nova," "National Geographic," and "The Ascent of Man" were mentioned as adult interests, while "Mr. Wizard" was mentioned by a few as child interests. Many of the participants had gone to natural history and science museums, but for most the experience was not particularly memorable. The importance of print and the relative lack of experience with other informal science resources may reflect the fact that they were unavailable to many of the respondents in the era they grew up. The older respondents did not have TV as children; scientists from anywhere but the major metropolitan areas did not have access to museums; and science-related clubs were few.

For obvious reasons, evidence of direct NSF influence was rare. In only one case did we find evidence of such a direct effect: one scientist had attended an NSF-funded Student Science Training Program during high school. Others noted less direct effects--for example, the influence of science television shows.

In general, there were few differences between minority and nonminority respondents or between women and men in terms of the factors most strongly affecting their career choices. As we noted above, a number of minorities and women in our sample faced barriers to the pursuit of science-related careers that white males did not face. In general, these involved a lack of support from teachers and other professionals in their schools. In spite of the barriers, however, women and minorities in our sample achieved at the same rate as their white male counterparts and were affected by the same factors. In all cases, the support of parents and other family members was the most important single factor.

Lessons Learned About Retrospective Assessments

The degree to which our findings are compatible with related research (e.g., Gustin, 1985; Sosniak, 1985; Zuckerman, 1977) suggests that there is some validity and stability to the evidence gathered by means of a retrospective study. Not only are the generalizations we draw similar to those drawn by others, but a comparison of our individual interviews with theirs presents remarkably similar data. The fact that even a small study like ours could begin to approximate the findings of more detailed work suggests that the retrospective method may be both efficient and informative in learning more about how people develop scientific interests.

Our limited effort points to several improvements in research strategy if such a study were to be pursued in the future. First, the younger the respondents, the better information one can collect. We found that younger respondents were able to identify better the sources of influence on their career choices because the crucial experiences were less remote. Older scientists and engineers were more likely to point to general influences ("My parents were supportive"; "I always did well in school"), but had trouble identifying specific experiences. Second, the use of contrasting groups might prove helpful. A number of the characteristics common to our group of scientists and engineers--e.g., supportive parents, excellent academic backgrounds--are probably shared by successful professionals in nonscience fields. It might have been useful to contrast our findings with those from interviews with a group of successful individuals in another field. For example, we might have compared advanced graduate students in physics and business. Third, individually focused retrospective studies should be used in conjunction with other forms of assessment. As is true of most forms of assessment, retrospective research focused on individuals cannot by itself provide a detailed explanation of the relationship between individual outcomes and science education resources of the kind NSF funds. The strength of retrospective studies lies in their ability to complement other forms of assessment.

Given these limitations, this investigation illustrated some of the benefits of looking retrospectively at the type of investment NSF regularly makes. Carried out with sufficient rigor and with a strategically chosen sample, this kind of assessment could indicate a great deal about what kinds of experiences shape individuals' interest in science. Evidence regarding the residual effects of the kinds of activities funded by NSF may be able to point to areas of high impact for future NSF investments.

The kinds of findings just described are useful to NSF planners in several ways. These interpretations are, of course, broad and not specific to particular NSF initiatives, but they can suggest directions for investment or shed light on the assumptions underlying particular initiatives.

First, the findings describe one aspect of the informal science education investment domain, by suggesting salient factors that play a role in the development of science interests and occupational pursuits. In this sense, they help one to understand, at a gross level, the phenomenon of informal science learning in a particular segment of the population.

Second, they put the influence of NSF-supported informal learning experiences in some perspective: the most powerful forces in these individuals' career development (to the extent those forces can be detected from retrospective reconstructions) are families and role models. Yet at the same time a variety of informal learning experiences, many of which are (or could be) among those supported by NSF, do play an apparently important supportive role in the science career development process.

Third, the findings confirm that NSF-sponsored experiences can have an appreciable influence that remains salient to individuals over time. Although only a few of our informants volunteered such information, the fact that they did so with little prompting suggests that experiences provided by television and student enrichment programs leave their mark on the formation of long-term interests and identities. It is also noteworthy that *any* of our respondents made note of these effects, when one considers that a third of the sample were too old to have interacted directly with NSF-supported resources. (To be sure, the findings in such a small sample could be coincidental; only research with larger samples could confirm or refute this hypothesis.)

Fourth, the findings point to provocative differences, as well as common themes, among respondents. For example, the fact that women engaged so little in tinkering yet so extensively in book-related science learning could suggest new targets for NSF investment intended to maximize participation among underrepresented groups (e.g., science hobby development in girls' recreational associations, children's tradebook offerings that maximize their appeal to female scientific interests).

Finally, these kinds of findings can offer some insights into processes that take many years to unfold. Although these can never replace the more meticulous account of career development over time that emerges from longitudinal research, this kind of study can help NSF to consider macro-level issues in current planning and program justification.

REFERENCES

Crane, V. (1987). *An exploratory study of "3-2-1 Contact!" viewership.* Chestnut Hill, MA: Research Communications, Ltd.

Elmore, R. (1980). *Complexity and control. What legislators and administrators can do about implementing public policy.* Washington, DC: The National Institute of Education.

Feynman, R. P. (1985). *Surely you're joking, Mr. Feynman!* New York: Norton.

Gustin, W. C. (1985). The development of exceptional research mathematicians. In B. S. Bloom (Ed.), *Developing talent in young people.* New York: Ballantine.

National Science Foundation. (1987). *National Science Foundation support for significant advances in fundamental engineering research as shown in publication acknowledgments 1965-1985.* Washington, DC: Program Evaluation Staff, National Science Foundation.

Sosniak, L. A. (1985). Becoming an outstanding research neurologist. In B. S. Bloom (Ed.), *Developing talent in young people.* New York: Ballantine.

Zuckerman, H. (1977). *Scientific elite: Nobel Laureates in the United States.* New York: Free Press.

Exhibit VII-1

THE INTERVIEW PROTOCOL USED IN THE FEASIBILITY STUDY

PURPOSE OF THE INTERVIEW: To discover influences and experiences that helped to form current attitude/disposition/involvement with science, mathematics, and technology.

APPROACH OF THE INTERVIEW: Start with a general exploration of what has influenced the interviewee and move toward more specifics. In particular, ask about informal science education resources only as a subset of all kinds of influences that may have been important.

(1) *Present occupation/interests:* What is present occupation? Professional interests? What are current avocations? Interests? (We are interested in understanding what influenced you as you grew up to develop these interests...). How would you describe your current interest in and/or involvement with things that involve science or math or technology?

(2) *School experiences:* Describe area of graduate study, college major, and high school interests. What kinds of things were you successful at? Not successful at? What did you think you could/would do for a living? Any teachers/professors that had influence on you? (positive or negative) Memory of math/science courses? Last one taken? Turned off by any? Any important experiences/critical moments you can remember? Memory of science courses? Do you remember SAT scores? Others compared to you in math and science? Empowered/made confident by science experiences? How did you know what you were good at? Overall, what influence did school have on your interest/involvement/confidence in science/math/tech?

(3) *Parents:* What do/did parents do? What kinds of things were they interested in? Was the environment intellectual? Active in hobbies? What hobbies/cultural things were they involved in? What did they share with you? (e.g., sports, music, nature, birdwatching...) What was their attitude toward/about science/scientists? Math/mathematicians? Engineers? Other technical things? What messages did they give to you about school? What was the message about "free time"? What message did they give you about doing/studying science/technical things? Empower you to action? Overall, what influence did parents have on your interest in/involvement with/confidence in science/math/tech?

(4) *Other influential people:* Who were role models to you? Who encouraged you? To do what kinds of things? Overall, what influence did these individuals have on your interest in/involvement in/confidence in science/math/tech?

Exhibit VII-1 (Concluded)

(5) *Hobbies, clubs, sports, music that influenced you:* What activities were you involved with as a child? Boy Scouts/Girl Scouts? Collect things? Fix things, tinker? Messing about with boats? Explore, learn about natural things? (e.g., birdwatching) Sierra Club or other environmental group activities? Overall, what influence did these activities have on your interest/ involvement/confidence in science/math/tech?

(5) *Readings that influenced you?* Library use? Magazines (related to hobbies/interests?) Particular books. News... Overall, what influence did your own reading have on your interest/involvement/confidence in science/math/tech?

(6) *Television or radio that influenced you?* Shows like "Nova," "National Geographic," "Mr. Wizard"? "3-2-1 Contact!"? Negative influences? Overall, what influence did TV have on your interest/involvement/ confidence in science/math/tech? What records or radio had an influence on you?

(7) *Cultural institutions?* Did you go to libraries? Museums? (positive or negative) Science museums in particular? Aquaria? Zoos? Parks? Overall, what influence did visits to these places have on your interest/involvement/confidence in science/math/tech?

(8) *Topics of science:* How interested were you in particular topics, for example:

Space
Environment/conservation
Astronomy/weather
Health
Computers
Other?

(9) *Current attitudes toward science/math/technology:* What is your attitude toward people who do science or math? Are they very different from yourself? "Understanding science and math is/is not something I am good at..." What do you think most influenced your current attitudes toward/involvement with/confidence about science and math?

Appendix

EXPLORING THE ASSESSMENT OF LEARNING IN INFORMAL SCIENCE EDUCATION SETTINGS-- RECONSTRUCTED DIALOGUE FROM THE MEETING

In this appendix, we present an approximation of the transcript from the meeting described in Section V, which explored the assessment of learning in informal science settings. For this meeting we brought together experts deliberately chosen to represent diverse fields and even more importantly diverse perspectives on assessment. The participants were:

Philip Morrison--Physicist, science educator, author.

Phyllis Morrison--Elementary science specialist, film and book reviewer.

Jon Miller--Political scientist, specialist in scientific literacy.

Elsa Feher--Physicist, specialist in cognitive studies in science museums.

Bret Waller--Art museum administrator, museum educator, art historian.

Vito Perrone--Evaluator, expert in inquiry-based science learning.

Keith Mielke--Applied research specialist in children's television.

Valerie Crane--Communications and marketing researcher.

Roger S. Miles--Exhibit designer and researcher.

Michael Templeton--NSF Program Officer, Informal Science Education.

George Tressel--NSF Division Director, Materials Development, Research and Informal Science Education.

Mark St. John--SRI researcher.

Patrick Shields--SRI researcher.

The dialogue that follows presents our reconstruction of the transcript of the meeting. We have tried to preserve the essential ideas and meanings as they were presented in the meeting, but we have edited and resequenced the quotes presented below. Thus, the content of this report is ultimately our formulation, not that of the participants. Our reconstruction has been reviewed by meeting participants, and as far as we can tell is a close approximation of the meeting discussion.

Mark St. John, Moderator: Let me pose the following question to focus our discussion: "Given the nature of the informal resources that are funded and the diverse ways in which people interact with them, what assessment approaches and procedures might help us learn about what people take away from their informal science experiences?"

WHAT IS NSF'S MOTIVATION FOR DOING SUCH ASSESSMENT?

Vito Perrone: To start out, let me ask a question. Is it fair to ask why NSF has to try to assess "learning" beyond, let's say, documenting what people do? Is there any need for NSF to go beyond that, to go to the level of what individuals learn, for example?

George Tressel: Yes. But whether it is a hard need or a wish is the question. The desire to do evaluation is both internally generated--because we're all intellectuals and we want to justify our own existence--and externally generated because Congress asks: "What are you spending this money for? Prove it is well spent." As a result, there is a strong desire for some measure of what's going on. In the area of public understanding of science, this pressure has gradually relaxed. When we first did "3-2-1 Contact!," it was a major issue. "You are going to spend millions of dollars and how will you know anything happened?" What I argued then, and have kept on arguing, is, first, that assessment must be broadly defined, and second, it must be feasible.

I've argued, first of all, that while we might wish we could see the behavioral change in kids and see some real change, we don't have the tools for it. We don't have the methodology to measure what happens to kids over the long term. We will be doing very well if we can just observe what happens and document that.... So, we are now gradually accumulating descriptive documentation.

Vito Perrone: The reason I raised the question is that I wanted to get a sense of whether you are operating in a mode where you are free to do assessment in ways that you feel are appropriate or...are you feeling constrained, under a pressure of a sort that Title I and a lot of other federal education programs seem to feel...a pressure to do evaluation in a way that seems to trivialize the learning process--where real contact with individual children is lost...and at that point I think evaluation can become one-dimensional. It may meet the requirements but it's not very useful.

George Tressel: Ironically, most of the pressure or the tendency to trivialize has come from inside the field instead from outside it. Some of the evaluators working in the informal field come with a preset methodology.

Vito Perrone: But the reason I've raised this question is that I'm wanting to know whether NSF, in doing assessment, wants to remain consistent with some of the philosophic orientation of informal learning or if it's going to have to alter it through its assessment mandates, and so I'm satisfied to hear the response that it's within the framework that we're talking about. Then I think you can talk about assessment without feeling the pressure of enforced artificialness. I even think a lot of the Title I evaluation is artificial. I don't like the designs, the RFPs, the way in which success has been defined *a priori*. I mean, it distorts--it distorts reading, it distorts natural meanings of reading, and it distorts some of what schools do. So my question was, "Do we have to operate within that kind of outside- imposed evaluation frame?"

Mark St. John: I think that is what we are here to do--to map out what kinds of assessment approaches one could do without distorting the nature of informal learning. We don't want the observation to change the observed. We want to find out what assessment methods provide too distorted a view so that we can avoid them. So we're looking for those kinds of guidelines here. And it's also important for us to articulate a clear rationale as to why we should or should not try different assessment approaches.

Phil Morrison: I just wanted to say that I very much like the notion you described of seeing "acculturation" as a goal for science education. This helps us to broaden our thinking about what we are doing. It is true that science and technology comprise a major subculture and are increasingly evident in our mainstream culture. It is important that more than just scientists and engineers be acculturated.... A second, much larger tier of the populace must be able to think scientifically as well.... And perhaps in our society it is only the subculture of sports that is equally, or even more, pervasive than that of science. It might be very interesting for NSF to study sports in order to better understand how people become a part of a subculture. After all, scientists, like professional athletes, ultimately need a knowledgeable and appreciative audience to play to, and in sports there is such an audience. Kids grow up playing sports; they are discussed in the family; they are on television; there are sports sections in the paper.... A multitude of informal processes throughout life combine to generate a "sports literate" population. To continue the analogy, there is a base upon which formal sports training at the high school level, college level, and professional level can build. There is a huge pool of potential athletes.... I think this is a useful parallel for thinking about the development of science literacy...."

DIFFICULTIES WITH THE "INPUT-OUTPUT" MODEL

Mark St. John: Let me raise the issue of a whole class of evaluations that are basically designed to make immediate and direct measures of the outcomes of interventions. They are of the type that says: we measure a certain characteristic of the input stream and then we walk around to the output stream and make another measurement, and then we subtract one from the other and we get a difference, and that difference is success. I mean, there is a whole collection of methods that are part of this evaluation approach. Some of those are profound, strong, necessary methods. But for the purposes of assessing the informal domain, to say we are going to measure gains--that we're going to measure cognitive gain, we are going to measure content gain, and attitude gains--I wonder about the appropriateness of this approach for this domain.

Patrick Shields: So you are saying there are "input-output" or "treatment-effect" studies...and the question is: "Is that a reasonable way to go?" And with a given input-output study you could span either a single interaction with an exhibit, a whole museum visit, or even several years, I suppose. I think this is the most common form of evaluation methodology and it is the first thing that evaluators tend to reach for.

George Tressel: I'll make up an imaginary example but it's pretty close to the real thing. If I do a study of "3-2-1 Contact!," I test kids beforehand to see what they know about how airplanes fly. Then I test them afterwards on the definition of pressure and things like that, and let's say that I can't see any change in their scores. So I report that there was no cognitive gain from this show.

What is worse, this kind of thing will be published because it is going to be "rigorous." It is going to have nice data that says before and afterwards nothing changed...and it won't be that the kids didn't learn anything...it will be that nothing comes out of the tests that were given.... And it may even become the conventional wisdom that "kids don't learn from '3-2-1 Contact!'"

To be perfectly honest about it, I worry about this kind of testing all the time. But that doesn't mean we don't want to know what goes on. So, let me display a strong bias about what's useful and what's not. The stuff that is most "rigorous," that follows the methodology blindly, is least useful. If you want "rigor," then you are likely to trade off utility for triviality. High rigor, high triviality. And worse, the most useful studies have no place to be published, because they tend to be anecdotal or semi-anecdotal in nature.

Mark St. John: I think you are looking at evaluation practice in a rear-view mirror. If you hang around the evaluators and people that do assessment these days, the fact is that a transition has been made by the field: it used to be the renegades who talked about qualitative methods.... Now it's quite switched around. In fact, the quantitative, experimental, rigorous guys are in hiding.... You have to remember that all assessment approaches are imported. There aren't any measurement approaches that are native to informal science education. There are no homegrown tools.... There are, in fact, warring schools of thought...your kind of Skinnerian scenario reflected the thinking of educational psychology in earlier years.... In those days it was clear--just get your objectives well stated and then measure them...that is the way to get progress...but I think that's kind of a past thing and now....

Michael Templeton, NSF Program Officer: It's not past in museum evaluation, it is the most common tool used--even in formative evaluation.

George Tressel: I can think of a study, done only a year and a half ago (by somebody who surely knows better), that asked questions as kids came in the door of the museum, asked questions as they went out, and then reported that they didn't know any more when they left than when they came in. I think, "Oh, how could you do that?"

Valerie Crane: But I don't think the danger in those models of evaluation comes in the fact that they are rigorous, but more in the assumptions they are making about the fact that you can capture informal learning using a very structured way of measuring the phenomenon. I mean, they are trying to measure learning by saying: "Here's a set of knowledge and by the time you go out the door you should have it."

So it's the assumptions of the people doing the research--they are off base. There is nothing wrong with their quantitative approach or anything else--it's just that they are really making value-laden decisions about what the expectations are... and about the fact that the learning experience has to be sequential, cumulative, and equal for everybody.

Keith Mielke: My feeling is that usually when this kind of evaluation happens, it's not so much because of a value decision that was made, but because we tend to measure what's easy to measure. We do what is most convenient...and what we know how to do.

LEARNING LESSONS OR PROVING RESULTS?

Vito Perrone: Would you review again the rationale for why we want to acquire evidence of individual learning in the first place? Are we trying to get proof of program performance or are we doing something else--what are we looking for?

Mark St. John: This is always a question that arises in evaluation--evaluation for what purpose? There are many functions of evaluation. One is program justification, internal and external. So, often that involves proof. But I think that is not the right approach--especially for this domain.... I think the tone we want to take is that evaluation is done to learn more about the enterprise. I don't think one should set out in this business to prove anything--in light of the difficulties in definition and measurement that we've heard.

Michael Templeton: Just a comment, a corollary to what you were saying--it sounds as if, if you really want to measure something in this domain effectively, the first thing you should do is avert your gaze so that you're not blinded by looking directly at what you think you should measure. Like looking at a dim star at night, you should look a little bit to the side. The worst combination may be looking at outcomes after talking with the instructional designer of a particular product and then attempting to make a connection between the two. That is a fatal combination; you are much better off doing these kinds of studies slightly misdirected. For example, say we put $200,000 into an exhibit. Now, you say, NSF should find out whether or not it worked. Wrong, wrong, wrong....

Mark St. John: So you should look at what exhibits as a class do or are capable of doing....

Michael Templeton: It is better to try to understand more about the ways that an exhibit functions, and how it functions in a particular setting, than to try ineffectually to find out whether or not we got our money's worth on a particular project.

Patrick Shields: So we use evaluation to build a rationale--but we do it in steps: we show that NSF can get good museums to take our money and to build exhibits. And we can find evidence that the exhibits are innovative ones. And then we're going to look for evidence out there that, in general, exhibits play a significant role in life. But we don't try to prove that this exhibit necessarily teaches X, Y, and Z.

George Tressel: The truth is, if you have a persuasive rationale for what you are doing, people will accept almost any kind of reasonable supportive evidence. Anecdotal evidence--videotapes of kids responding, the descriptive evidence on how many kids are watching it. One bit of evidence that says "kids and their parents go away and do something afterwards" is worth its weight in gold. It's worth all the evaluation effort just to be able to document that.

Keith Mielke: Let us say we will seek that evidence.

Valerie Crane: I feel the most fruitful areas for assessment are not found in trying to give report cards and grades. We ought to stop trying to fit projects into boxes. I think we should to get more into the processes of learning and move away from the boxes and the grades and into understanding more about how informal learning works. We have not spent much time talking about that here. We have to look at it much more in detail and in context. We have to understand what kinds of individuals come into the loop, what happens to them once they get there, what this experience is like for them--really tracking in a very careful way what is this process we're talking about, trying, as we go along, to define it.... I think evaluation should be much more of a process of looking at the details of what happened. We know lots and lots of things go into each person's experience, and I don't think we should be worrying about giving grades and doing assessment.

Mark St. John: I don't see the kinds of process studies you are describing as being independent of assessment. To me, that is assessment.

Valerie Crane: I just think it is important to stay away from things that are program scores--"boy, did we spend our money well" kinds of things.... I understand what motivates the desire to measure these things, but it could be very frightening, when we aren't even sure what our outcomes should be and we start leaping to models of efficiency.

WHAT CRITERIA OF EVIDENCE DO YOU USE? WHAT DO YOU TRUST?

Mark St. John: I want to explore another question here: What kinds of evidence are credible, valid, and useful? What would cause us to say that the investments in programs...the things that we spend our money on...were successful? What measures or criteria would we trust?

George Tressel: One thing would be self-reported anecdotes about projects. Anecdotal information can be very powerful because it includes gritty little stories and details so that you can see through some of the rhetoric to the reality.

Jon Miller: So you would trust it when the people you fund tell you stories that show they did a good job.... That sounds a little suspicious....

George Tressel: The people we fund usually think they did a good job. But, that is not all they will tell you...and when you get to hear the real grant recipients through the anecdotes they tell--when you get to hear the voices of the kids or the comments of the adults, then you can begin to draw your own conclusions. Probably the best kind of information is the report from a principal investigator who says, "God, we had a lot of trouble with this. Let me really tell you how and why we had trouble." When you begin to hear the pieces of a project that didn't work, balanced off against the pieces that did, then you begin to get a calibration to what else is said--a reality check.

Michael Templeton: We received an evaluation report recently that was very persuasive. It says, "We did this and some of the students said this and some of the students said that and we had a hard time getting such and such going... The leaders didn't know how to do a particular unit and we really had a rough time with it, etc." The value was that it was good news and bad news and not just saying, "Wow! Were we wonderful." Instead, it reported, "This is what happened." Boy, is that a persuasive report.

Vito Perrone: For you that's much more persuasive than if someone came in with some numbers, then.

George Tressel: We are looking for the "good news/bad news" combination. There is always good news and bad news. And if someone tells me only good news, I don't believe them. Neither does anybody I talk to.

Mark St. John: The goods news/bad news criterion for evaluation, I think, is a more specific example of putting assessment in the larger context of assessment aimed at learning: "Here's what we did, here's what we learned"--instead of "Here's my proof that what we did was good." So it puts evaluation in the context of "Here is an experiment we tried and here's what we learned from the experiment."

Keith Mielke: Yes, but we have raised but still not resolved an important issue: the persuasive power of the credible, resonating anecdote versus the non-status of the anecdote as scientific evidence.

Michael Templeton: Let me give you another example that is more concrete--an out-of-school education project. The reason we funded that project was because we thought that the model had some possibilities for propagating elsewhere. And it was a stick-your-neck-out award because it was a local project. The project is now nearing completion. The principal investigator responsible for the project came back and said, "I'd like a small supplement so that I can travel around the country and hold workshops with a dozen different groups that want to start their own version of our project. I think that this is going to lead somewhere." That alone is an evaluative statement: here's somebody investing their own energy and time in what we had hoped would be the next step--to take something that did work and find ways of disseminating it--without a lot of prompting from NSF.

Jon Miller: But how do you know that there is any value to what this person is spreading? How do you know that anyone is learning anything?

Michael Templeton: The main issue in that project was not whether or not you can teach children science with volunteers. That's not too hard to figure out. The main issue in that project was whether you can construct an innovation that has legs so that it spreads on its own.

Jon Miller: Let me argue that question. You know very well that I like the programs that you are doing, so let me just be harsh for a minute. I think this kind of anecdotal assessment and the kind of approach that you are talking about may be exactly the kind of thing that gets science education in trouble.... What happens when they ask what your evidence is? In this case I don't think you know anything. We are trying to think about doing evaluation the same way we do science...and it seems to me that you are doing nonscience when you ask those who have a certain enthusiasm for their project--and they see it through those glasses--and they come back and tell you what they think happened--I suggest you are getting a very biased set of views. To say that a project has legs on it, that responds to the interests of those doing the projects, not to the interests of those who are on the receiving end of the educational effort.... I would argue that on the basis of those things just described, you have no evidence whatsoever that the program worked.

Michael Templeton: I never claimed to have carried out an assessment. And I don't think I was asked whether I could carry out an assessment. I was asked what I considered to be evidence of success.

Valerie Crane: No, it was asked whether you considered the project to be successful. And I think what he was saying is that you don't have evidence that it was successful.

George Tressel: Our view of success depends on the fact that we are in the investment business. One thing that an investor must worry about is whether or not the audience likes the program. Another one is the strategy of investment--does the project grow "legs"? Does it get up and walk someplace else and reproduce itself? Because impact for us is the most critical thing. A local program might be wonderful, but if it won't get up and "walk," then NSF doesn't belong in it.

Jon Miller: But to ask the question about having legs--it's like stockbrokers thinking they are doing well if they have a high rate of turnover in their stocks.... There is a difference between volume and return on investment. The real assessment issue is whether you made any money.

Mark St. John: It strikes me that we are arguing at different levels...it seems necessary that the rationale for a project must hold up at these different levels of detail. Failure on any level kills the whole thing. That is, if a local project remains local, then from NSF's investment point of view it would be a failure no matter how much people learn from it. If the project, on the other hand, gets propagated nationwide, it would be a failure if the kids didn't learn anything, if the materials were bad, if it was sexist--whatever. That is why we're having this meeting at the micro level--one has to make an argument at many levels to validate the success of an initiative, to support the rationale for investing in these things.

SUMMATIVE VERSUS FORMATIVE EVALUATION

Mark St. John: We have talked about standards of evidence and the different levels at which initiatives can be assessed. Let me ask about purposes of evaluations. What about the formative/summative distinction?

Keith Mielke: I look upon the distinction as a function of use and not a design issue. I mean, if you use the information formatively, then it's formative. If you use something summatively, it's summative. In our case, when we use information from the evaluation of the end of a TV program series to shape the next series--say, in "Square One," you know we're going to have a lot more game show activity in the second season of "Square One TV" because they work gangbusters--we know that in a quick and dirty summative way, if you will--and so we're using that information formatively to help shape the way we do it the next time around. And then it can be sent to funders for whatever decisions, or to inform them about the series that they would want. So we do use them both ways, yes.

George Tressel: "Summative evaluation" to people usually means "tell me whether or not this was a good project." That doesn't mean anything to me. We're trying to find out how to make something happen, and in that context there is no "bad" project. If it doesn't work, we want to know it, and we want to know why so we don't do it again or so we can figure out what else to do. It's all formative evaluation from our point of view...it's future oriented and that's why Valerie's "3-2-1" study is so important to us. It's not a question of "was this a good project or wasn't this a good project"... but it tells us the good news and the bad news so that we are in a better position for the next stage. The whole business of saying "this was a good project" or "this was a bad project" is wrong because nothing in this world is so simple.

THE DESIGNER'S AND THE CONSUMER'S PERSPECTIVES

Mark St. John: Let me raise the same issue another way. There perhaps is another distinction that may be useful here...and that is the perspective or frame of reference from which judgments are made.... There is the perspective of the designer and then there is the perspective of the consumer: those who use the informal science resources.... An analogy to the evaluation of consumer products can make this difference in perspectives clear. The magazine *Consumer Reports* doesn't particularly care what the GM or Chrysler designers had in mind. And that's a good thing. It only worries about factors that are important to consumers--its criteria are derived from consumer needs and interests--their safety, their comfort, how often you have to fix the car, and so forth. Some of those may overlap with the designers' criteria, but they are not always the same.... This is why drug testing uses a double-blind design.... It also is why Scriven pushes the notion of "goal-free" evaluation--it doesn't really matter what the drug manufacturer or the car designer or even the exhibit designer had in mind. In all of these cases, the consumer does not care.... Perhaps this distinction of perspectives may be useful in helping us formulate general guidelines for evaluation in this domain.

Valerie Crane: Well, doesn't the manufacturer know some of the consumer's interests in the very beginning?

Mark St. John: Ideally, yes. But then you see that there are Edsels around. It's not always the case.

Keith Mielke: Mike said earlier that there are severe limits in answering the question, "Did you get your money's worth?", and there are problems in using an enthusiastic producer's version of what ought to be the index of success. Mike is basically making a "goal-free evaluation" appeal, without grappling with all the downside of that particular position. Valerie Crane also makes what to me is the important point: the producer's assumptions about the target audience are informed assumptions if the formative research is functioning properly. To assume that producers and consumers are completely independent is to assume the nonfunctioning of the formative process. In this light, I don't think that Mark's observation that such a system is not 100% error-free (Edsels do occur from time to time) is a powerful argument.

Elsa Feher: I think there are two questions one could ask. In some general sense, you can ask, "What happened?" The other question is: "Did what I wanted to see happen actually happen?" Now, those are very different questions--one looks at the experience from the provider of the experience; the other looks at the receiver of the experience. Now, if you are the provider of the experience and you *a priori* set down what you are going to look for, then you assess the learning experience with respect to those things.... If you look at it or attempt to look at it with the eyes of the receiver (the consumer), then you have to ask a much more open-ended question. Then you try to put yourself into the world that the visitor or child has constructed and you interpret what the child does and says in the context of the child's world--rather than your world. I think that puts a very different perspective on the evaluation.

Keith Mielke: I agree that one should be open to any form of impact, through questions such as "What did you get out of this?" rather than to acknowledge an effect only if it happens to fall within an *a priori* checklist of anticipated or hoped-for outcomes. In this perspective, the goal-free enthusiasts seem less dangerous than when they assume that the producer's knowledge and expectations are irrelevant.

Vito Perrone: You can ask kids what they took away from the program and it's possible to do a qualitative review of those kinds of responses. I think one can make judgments about the quality of what they say. I think it is possible to do this, and if you do it, it shows that you take your audience seriously. I think you can learn a lot from this approach.

Phyllis Morrison: We did this kind of thing at the children's museum...and it was terribly useful--it was useful to us right there and then.... I would simply stop kids as they left the museum and say, "Tell me about what you have been doing." And essentially what I was listening for were the things I hadn't heard before, things I didn't know about--I was looking for things that were surprising to me. And I got a lot. You have to ask and then listen very hard, and when you do that, you hear some very surprising things. And those surprises were very good to hear.

Mark St. John: So it's a different focus. Instead of asking, "Did you learn X?" or "Did you learn Y?" or "Did you learn Z?"--it turns it the other way around. And somehow this seems more appropriate to the informal domain since the kids were self-motivated and self-guided in their learning anyway.... Perhaps in the informal domain what the designer had in mind, for all of his or her cleverness and good intentions, may play a small role in determining what ends up happening in that setting--in research language, we could say that the designer may at best have control over very little of the rather large variance that we see in such settings.... So what the interviewer is saying, in effect, is, "You, the visitor, are the one who decided what you did and what you learned; now tell me about it--whatever it was. Tell me about what you saw, what surprised you."

Elsa Feher: But I want to point out that it is also true that this visitor-centered approach can be very useful to me--as the designer. And as an involved and invested designer, with my own lofty goals, it may be the only way I can do evaluation that doesn't limit my perspective or put a filter on it. And, for me, in practice this approach of asking kids what they have done, what they have seen, has really worked very nicely. It has been very useful.

Bret Waller: I want to reiterate what you were saying--one way to measure success would be: "Did we accomplish what we set out to accomplish?" And at first I thought that sounded perfectly good and reasonable--but then I thought--is a book always about what the author intended? If it isn't, does that mean the book failed? What if you found that you didn't get what you wanted to, but that you did get something else? You have to judge: well, is that something else good? And in order to say that, you've got to have been able to make some value judgments.

Valerie Crane: But I want to argue that that is not the job of research...it can never tell you what the value of something is--it can only can tell you that this is what happened, that this happened more than this, or this happened less than that.

SPECIFIC GUIDELINES AND SUGGESTIONS FOR NSF'S APPROACH TO THE ASSESSMENT OF INFORMAL LEARNING

Mark St. John: Let me propose that we might think of two levels of assessment. One is large scale and takes a broad view: assessing the informal science investments of the Foundation. The other level is assessment done within the projects themselves. One idea we want to dispel right away is the notion that the assessment of NSF's informal investments is the sum of the evaluation done by the projects. It is not true that if you collect all of the final project reports, and put them all together, then you have done an assessment of the overall initiative. I think that another whole perspective is needed.... NSF has meta-questions that inherently extend beyond the aggregate of the project specific evaluations. But let us first take up the issue of doing project-level evaluation.

WHAT APPROACH TO EVALUATION IS USEFUL AT THE PROJECT LEVEL?

Roger Miles: There was a period when instructional designers were enjoined to write lists of objectives intended to describe accurately, and in detail, exactly what the educational transaction was supposed to consist of. There is a flaw in this whole approach--which is fairly deep--and that is the assumption that you can prescribe with a list of any sort interactions and transactions with informal educational materials. In the field of exhibit evaluation, I would say that to construct such lists and use them to design testing procedures and make changes in exhibits has not been especially useful.

Elsa Feher: What these objectives take the form of is like this: "90% of the visitors will have spent at least 5 minutes manipulating this exhibit." Certainly it is true that unless someone uses the exhibit for at least 5 minutes it is unlikely that they will get anything out of it.... So I would say this kind of objective states a necessary condition that is in no way sufficient.

Roger Miles: But the problem is worse than that. The phrasing of the behavioral objective leads to semantic problems with designers. When you start writing that 80% of the visitors will do this and then will be able to answer these questions to this degree of accuracy under these conditions, there's no real point of contact with what actually goes on in the informal education environment; and designers know this, so in order to get off the hook they start prefacing their objectives with phrases such as, "under ideal conditions...," which simply undermines the whole approach.

Keith Mielke: I think there is almost a disservice that's been done by this kind of criteria setting because what ends up happening is that, with most exhibits, you just have to throw your hands up. You can't think this way--not only does the viewer or visitor not behave that way, the instructional designer doesn't behave that way either...they're totally incapable of designing an exhibit based on the idea that 80% of the people should push a lever. That's not the way the design works. Maybe that is why there has been a lot of resistance to incorporating the results of museum research--because some of the research models that are out there are so off-base.

Jon Miller: But surely you do have goals for the visitor. If you expect visitors to improve their understanding of A, B, C, or D, why not expect 80% to conform to some criteria? Surely you must have in mind when you start some areas in which you expect knowledge to improve. Perhaps it is the false sense of precision that you are reacting to.

Keith Mielke: It is also the exact specification of the nature of the transaction with the exhibit that I find incompatible with the informal environment.

Roger Miles: What we do in the early planning stage is to list some broad aims for the exhibition. Then, during the design phase, we tease them out into detailed lists of teaching points, stating clearly what we want to get over. But, it doesn't help to wrap things up in the language of behavioral objectives.

Elsa Feher: Let me describe how we approach formative or front-end evaluation.... What we started with was this naturalistic observation, watching the thing and tinkering with it on all kinds of levels and watching what the people did with it, and then at a certain point we felt we had an exhibit that did what we wanted it to do. So then we had interviews with selected visitors--which is also called "cued interviewing" and we have called facilitative interviews--where we asked the people to do certain tasks. These are all very manipulative, very interactive exhibits, and we asked people to do certain tasks at the site of the exhibit. Now we knew what we wanted as an outcome of those tasks, and in facilitating them we were trying to see the potential of the exhibit...we wanted to see whether in fact it was possible to get from that exhibit as much as we wanted to get from it. Was it possible to lead these people to have at that exhibit the experiences that we wanted them to have. And it is not always possible to do. Sometimes it turns out we wanted them to see certain things and no matter how hard we tried they weren't seeing them. There was just no way. Therefore, it was unreasonable to have those expectations--if they couldn't do it when we facilitated their interaction, for certain they wouldn't be able to do it when we were not facilitating them.... So what this brought us to was a rewording, if you wish, of our objectives--one that was much more visitor based.... In essence, we were using both the designer and consumer perspectives.... They are not antagonistic. They are not even separate. You start with one and you end up with the other. So we found you have to work at it from both ends: you have to use evaluation to bring together the visitor's perspective and the evaluator's perspective.

Roger Miles: I mentioned yesterday an example of summative evaluation we had done with a big exhibition called "Discovering Mammals." We actually did a lot of front-end analysis at the beginning, i.e., at the planning stage. It might be interesting just to mention some of the things we attempted. We wanted to know what visitors brought with them to the museum. Did they know what a mammal was? Could they distinguish mammals from nonmammals intuitively? We wanted to know which groups of mammals they were particularly interested in--but not to let them tell us what we should deal with. We found they weren't very interested in certain groups, or in certain topics in mammalian biology. This is an exhibition about large mammals and yet they're not interested in the problem of size, which is very crucial in understanding the whole animal kingdom. So it told us we would have to work hard to connect some of our messages with people's interests.

All of this front-end research is directed at getting the whole design process off on the right foot, because there's no point in taking bad designs forward to formative evaluation. I do believe resources for empirical work are best spent as early as possible in the whole process of planning, designing, and producing exhibits.

Keith Mielke: You can add sort of a footnote to that. I think it's far better to put the money into empirical work either during the design of an exhibition or even before that stage...into audience analysis or market research--whatever you like to call it.

Michael Templeton: That's a pragmatic issue in the program because we have to work fairly hard to get people to instill evaluative components in the projects. It's the area about which even an experienced practitioner is naive. We have made a conscious decision in the program to try to convince people to develop from the beginning diagnostic--formative, front-end, market testing--strategies simply to make sure that the project has the best chance of internal success.

So when somebody asks me, "I'm going to spend $10 on evaluation; how do you think I ought to spend it?" I reply, "You probably ought to spend $7 or $8 up front. And when you are through, if you want to find out something at the end, spend $1 or $2 on summative evaluation.

Valerie Crane: But only once the project is completed can you learn how it is being actually used: what is the real sequence of events that happen as part of the total learning experience? This is something you can only get once it's finished. But you are right--how you allocate dollars relative to one or the other kinds of evaluation is another question.

Keith Mielke: In the rules of classical research, one must state up front the intentions and expectations, and then not deviate from them until after all data are analyzed, because to do so is to ruin the integrity of the research design. The evaluation of informal learning projects, however, should not preclude whatever midstream adaptations are called for by the available feedback.

Project development and production are processes that are inherently dynamic. Rationality demands that projects incorporate changing conditions, as well as increasing levels of being informed, into the thinking and strategies over time. Back at the first round, therefore, project leadership is understandably reluctant to "bet the farm" on a specific, unchangeable goal structure, and concurrently to be required to ignore all the experiences and insights gained in the production and formative research process. At any single point in time, there is--in theory, at least--some "best assessment" or "best snapshot" that can be made, but this is subject to change over time.

Mark St. John: So development shouldn't be preordained. I mean, you don't have to set the evaluation up years ahead of time, and stick to it rigorously. It's not a controlled experiment.

Elsa Feher: That's right--you are not in a position to have a controlled experiment because you are shaping it as you go.

Keith Mielke: I think it is very important to remember that one of the defining characteristics of informal education is the enormous variance in the audience that comes to it in the first place. We're talking about a research tradition that's not based in informal education. When we talk about specifying and measuring behavioral objectives, that is coming out of a tradition of homogeneity where you can say, "For this slice of instruction, with these assumptions, I would like this message to have this effect." But it's extraordinarily difficult to take something that is open to the full sweep of television--from the very young preschooler to the very old person watching the same program--to say that it makes any sense at all to talk about a behavioral objective guiding that enterprise.... One might, in principle, conceptualize a grid of objectives consisting of a very complex set of input characteristics, demographic clusters, and an equally complex set of outcomes. OK, I acknowledge that I have yet to actually master such a technique, but the principle is still worth holding up for consideration: one does not have to have a single set of goals that's supposed to apply to everyone; instead, it is in theory possible to conceptualize a widely variegated set of goals that do in fact anticipate great heterogeneity in the designer/consumer transaction. In television, for example, one could state goals for 3-year-olds that are completely different from the goals for 12-year-olds or 80-year-olds. Then we could do many subdivisions of that, such as by background knowledge levels, and so on.

I've tried to grapple with that, and in actuality it bogs down in its own weight--so it is very difficult to say up front in informal education that you're going to have these behavioral objectives and outcomes.

Mark St. John: But are there any general design principles? Is there research that we can do that will help us learn what criteria make for a good exhibit or a good television show? Do you have general design principles that you've tried to communicate?

Keith Mielke: When I was in graduate school, I really did believe that if we just could identify and control enough variables, we'd finally get them all into the corral.

BUILDING EVALUATION CAPABILITY

Michael Templeton: One of the rules or principles operating here seems to be that projects are heterogeneous. There is also a set of heterogeneous assessment techniques that are effective in understanding learning effectiveness in projects and useful in influencing the course of design of the project itself. These come under the rubric of front-end assessment strategies. They appear to be highly valuable to the projects themselves. Nobody has suggested that data generated for front-end purposes would not be useful for overall project evaluation...or that the data generated within the projects is not likely to be useful in NSF's overall assessment interests. One might want to assess the penetration of front-end evaluative strategies in projects as a way to get some sense of the sophistication of project design. The shared knowledge of those strategies and how they work and don't work would be very useful to the rest of the profession. From everything we've said, it is clear that at present they are inadequately shared. Informal education is still a field in which there is inadequate communication, not only evaluator to evaluator but project team to project team. There are many project teams in isolation from each other, learning mostly from within a project and not learning from other projects.... God forbid that two television production teams should actually talk to each other. Those are morals I would draw at the project level.

Keith Mielke: I agree that formative data can't serve a summative purpose, but the extent to which a proposal reflects formative sophistication may be a useful criterion for NSF funding in the first place. This should all be put into printed English. As an extension, perhaps the level of interproject sharing of information should become a funding criterion also, but that gets hairy. I know from past observations of federal funding how research procedures, when framed as funding prerequisites, can be responded to in letter but not in spirit.

Jon Miller: Maybe it's worthwhile for NSF to think about ways it might invest in the sharing of that knowledge, in the building of evaluation capacity at the project level. It might be worth looking for ways to continue to help the field develop its capacity for doing and sharing these techniques--maybe in PI meetings--but somehow to continue to look for ways to share those ideas, that knowledge and that expertise.

Michael Templeton: As a practical matter, reviewing proposals and making decisions about which projects to support--every project has its own vision, its own notion of what front-end evaluation means. You do not write proposals for proposers. You do not generate plans for people. You make suggestions, and you hint at various things and you judge them according to what they propose to do.

Mark St. John: But wait a minute.... You set all kinds of criteria that ought to be met in order for a proposal to be funded. There is no reason one couldn't set fairly severe criteria about formative evaluation if one wanted to.

Michael Templeton: Sure. If you set very severe criteria for formative evaluation you would fund few projects because formative evaluation and front-end design are the least well-articulated aspects of any project anywhere in the directorate. The number of proposers who are sophisticated and knowledgeable about front-end design is quite small. That is a characteristic of the field.

Valerie Crane: We went through this in television, and I think others will attest to this: that if you begin to make it somewhat of an expectation--I'm not an advocate of detailed formal guidelines--but if you at least require a front-end strategy as part of the proposal, I can promise you that magically people are going to find the resources to do better research... because if the money hangs on it....

Michael Templeton: We do that, and I think we have, in fact, had such an effect. This is an area in which the informal science education program has had rather persuasive leverage on the field itself. But most exhibits are not funded by NSF. Most museums don't keep evaluation and marketing research people on their staffs. Funding from federal agencies is the exception rather than the norm, unlike television, where a small number of sources provide most of the money. It is a different market environment. You don't do things by fiat; that doesn't work. Our goal is to build in an understanding of evaluation and to develop an expertise. You don't accomplish that by setting requirements. If you say to a proposer, you must have "type A" evaluation, you will get it. But it will be blind adherence to a formula. What really matters are the transactions inside the project itself. You want the exhibit fabricator talking with the person who has designed the survey form, the evaluator talking with the curator and the designer--and this cannot be mandated.

George Tressel: One thing that will help over the long term, I think, are successful models of projects, in which evaluative strategies are powerful and appropriate and are done in an ongoing way. Another thing that might help would be horror stories of things that went badly because there wasn't good front-end evaluation. Part of the problem in the profession is that you bury the mistakes and rarely share much public discussion of the things that started wrong and got worse. NSF can only have a limited influence in this area. In every discussion that I have with a proposer, I encourage formative evaluation.

INITIATIVE-LEVEL ASSESSMENT

Mark St. John: Let us switch to the larger picture--how can NSF assess its overall strategies for investing in the informal domain? Remember, we want to focus on the assessment of what individuals get out of the resources that are funded. What methods or approaches appear most promising? And rather than just getting an inventory of methodologies, it would be very helpful to have a few well-chosen words summarizing why each of these is particularly hitting the target of informal learning.

Longitudinal Assessment

Jon Miller: I think part of the assessment problem goes back to a broadly held perception in the academic community--and that is that evaluation should focus on the narrow study of one exhibit, one magazine, one TV show, and so forth. It seems to me that NSF and other funders should be willing to support projects that are larger in scope, more comprehensive, and multiyeared.... I think the kinds of questions that people ask in longitudinal studies are getting at appropriate kinds of questions for the cumulative nature of informal learning.... Most of the investment hypotheses... in both formal and informal education...are change hypotheses--that someone starting at some point in time changed in some fashion by the next point (or points) in time. And given the long-term nature of that change, a longitudinal study is the appropriate approach to exploring it.... If you try to do an experiment that is a cross section in time, it generally will show you what you are already suggesting--that is, it will show you no change.... And it also seems to me that we know that almost all science education and formal science education programs benefit some students and probably don't benefit other students. The real question is not does it have a benefit but who does it benefit and how does it benefit them? And why does it not benefit the ones it does not benefit? And to answer those kinds of macroquestions you need to have a broad-based longitudinal study.... And it seems to me we do know how to do that research. It's just that it's terribly expensive. It's not hundreds of thousands but it's millions....

Mark St. John: The point is that in informal science education you're really interested in changing people's attitudes and behaviors over the long haul. This process of acculturation is a long-term process--it takes time and while it may be easy to do a little input/output measurement thing right there in the museum that day--spend two minutes in front of the exhibit--that's not your goal. Your goal isn't to get them to know a little more about it...and so the whole new way you're going to study the process is to look at them for a long period of time.

Valerie Crane: Can I speak about the appropriateness of longitudinal studies for assessing informal learning...because in terms of trying to measure effects after a long period of time, the trade-offs are that the longer that period becomes, the less able you are to attribute any changes to anything having to do with informal learning. You're not going to know why people are changing: you don't know if it's because they went to Alice's exhibit or they watched Keith's television show or they went to London.

Keith Mielke: This is a crucial point about longitudinal studies: it is not possible to attribute causality to single interventions over long periods of time.

Experimental Studies

Valerie Crane: I also think that we haven't talked a lot about experimental studies. There are some good chances to do pretest and posttest kinds of things in the design phase of projects, and you can do fairly strict experiments when you have the opportunity to do them. In terms of the assessment question, the experimental method is a very powerful method when you have opportunity to do it.

Mark St. John: But what I heard yesterday was that fixed design, pre-post measurement, even in the affective domain, wasn't very convincing or useful to most of the people working in this domain...and now I hear that experiments can be useful.

Valerie Crane: What I think we were saying is that when you have a museum exhibit in a natural setting, then a rigorous experimental design is probably inappropriate for assessing the success of that exhibit.... But what the television production people have sometimes is a situation where they have a short clip, and you've got maybe 6 months before you go on the air with it; then you can set up a situation to see what the effect of watching the clip is.... It is like a "cued interview".... It gives you extraordinary opportunities to go out and do controlled experiments with different kinds of populations to see the effects of exposure and nonexposure--to do those kinds of things.

Jon Miller: I can understand using experiments to try and see the impact of various materials on different groups...but let's not forget that it is an experiment done in the studio outside of the public broadcast medium. So the experiment is only an approximation. It is not an experiment on the true use of the material. Once it is broadcast, you can't have an experiment any longer because now it's available to everybody--the audience is self-selected and you can't control the kind or degree of their exposure.

Mark St. John: So I still don't hear sentiment for a control group experiment in the natural setting.... Is that right?

Keith Mielke: My experience has been this: that in trying to capture a broad field of possible effects of informal learning such as television, any instrumentation that tries to presuppose the form of the effect is probably going to miss it. And that means the measurement needs to be open to any kind of expression. So that if I have on a checklist, "Did you buy a microscope?" the chances are very high that "3-2-1 Contact!" is not going to score a significant difference on it. If I go in and ask, "Is there anything that you've learned?" or if I ask the parent, "Has your child exhibited any behavior...?" the instrument can take anything whatsoever that pops up into visibility. You have to maximize the sensitivity of the instrument because these effects are frequently subtle.

Retrospective Studies

Mark St. John: Retrospective studies that take an "outside-in" approach can help provide a backwards trace of the chain of events in the initiative hypothesis.

Phil Morrison: I tried to think of a very successful and specific example of informal education, and I thought of George Gamow's *1-2-3 Infinity*. It was an extraordinary book, and I think for a very long time it had a major influence on the kids who decided they were going to study some science in college. It might be very interesting to go and talk to those kids who grew up with the book and get some anecdotal records of what it meant to them.

There is an immense need for that kind of thing--to document that books, museum activity, hobbies, etc.--all the things and activities that we deeply believe in--that these things have a significant influence on people...and I don't know of any formal efforts at documentation that try to do that.

Mark St. John: As part of our SRI study, we are carrying out a very small effort in that direction...we are interviewing scientists--that is, people who work in a wide range of science, engineering, and technical fields. And we are looking more specifically at a subset made up of women and minorities. And so, instead of asking how a particular exhibit affected them, we work at it from the outside in...we go at it backwards, asking what influenced the development of your attitudes and skills in science? How did your parents play into it? How did your schoolteachers influence you? How did science courses affect you? And then we ask, "By the way, did you happen to go to museums, did you happen to ever watch television, did you ever go to libraries, did you like reading science magazines or books? Did you have any hobbies?--all these kinds of things. Basically looking backwards, probing their experience, for evidence of "residues" of important informal influences.

Phyllis Morrison: That kind of effort might help to learn what questions to ask--in the study of the resources themselves.

Phil Morrison: Another interesting idea was brought up, but I don't know if it's feasible. Perhaps we could identify events in the past which appear salient...we know their time, we know their audience--or demography. Can we recover some signs of what impact those events had on people of that educational era? We should look for people who remain with that history in the right time of their lives to see if any residue was consciously left. It would be a very interesting study to make.

Mark St. John: Also, the work of Benjamin Bloom falls into this category--the chapters in there on the creation of research mathematicians--and similarly, Zuckerman's book on development of the Nobel prize winners...she describes well how to do the kinds of in-depth interviews that are needed.... She describes the methodology very well, but more importantly, she and Bloom point out the very consistent patterns of development and influence...this raises the hope that you might find very common kinds of patterns of how informal resources influence people.

Michael Templeton: Look, the value in this is not solely in trying to identify activities that somehow empowers, engenders, or breeds scientists. There is a general value to experiments of this kind to parse out those experiences that are responsible for what any adult knows about science. Knowing how and where an auto mechanic learned about astronomy is probably as important as learning where and how an astronomer learned about astronomy.

Mark St. John: Especially in the context of informal science education, I think you're right. And a subset of your broader view is the narrower question about when and where have events motivated individuals to pursue science careers.

George Tressel: But the problem is how to identify people who have a general interest in science. Your problem is that it is much easier to find a scientist who has interest in science than it is to find the auto mechanic who has an interest in science.

Michael Templeton: Perhaps you can find a thousand subscribers to *Scientific American* who are not members of the AAAS....

Phil Morrison: I agree that the more intriguing question was the study of how the larger population is influenced by informal science resources. I think those who are on the track to becoming scientists are already predisposed to seek out those resources. It is more interesting to understand how the general public works.

Jon Miller: I want to issue a caveat about this kind of research. We want to be cautious, not to simply go out and advertise to potential applicants that we are now looking for retrospective studies...because the method itself has a long way to go yet. It is something we ought to look at, but I don't think the technique is yet reliable enough to be of the same scale as other kinds of studies where there are reasonably known methods and reasonably sound statistical techniques. If you think of oral history, which has more of the anecdotal, humanistic flavor to it, people have thought that you could interview people and ask them to tell you what happened. This looked like a cheap way to get a longitudinal study. But there are substantial errors in what people recall.... It seems to me that what we ought to urge is some experimentation to see if the retrospective approach can be made useful.

Expert Opinion

Mark St. John: I wonder if there is a role for expert judgment in the evaluation of informal learning.... Eisner calls this approach educational connoisseurship or criticism.... I would argue that we rely on expert judgment in deciding which proposals to fund, so why not rely on them, at least in part, to judge how well the projects have done? We have some faith that there are experienced people around-- who can look at resources and judge the likeliness of these resources to foster learning. They could also watch how people interact with these resources and judge the educational quality of that interaction.... Since we don't trust the instruments or measures we have for directly assessing informal learning, shouldn't we at least add the judgment of experts to our assessment repertoire?

Keith Mielke: One problem I see--it's been my experience that those who are experts in science are certainly not experts, necessarily, in children, or experts in TV or anything else outside the very limited domain of their discipline. Sometimes in talking with these "experts," I find that the lowest person with whom they ever speak is a postdoctorate person, and then, when they give advice about children, they miss the mark wildly.

Mark St. John: But what about you? You're the kind of expert I had in mind. I would trust your judgment in looking at a new TV series because your expertise is hybrid in nature--you've been around scientists, you've been around children, and you've been around television. You're the one I had in mind, not necessarily the scientist.

Michael Templeton: I think you put more primacy to opinion than you should. The reason we use expert judgment in deciding which projects to fund is because we have no alternative. We have a set of decisions that have to be made. We use expert opinion, not because we think it's a perfect method, but because it works.

Keith Mielke: I'd like to elaborate a little more on the expert opinion notion because I still don't quite hear the ideas I have tried to articulate in my head. I believe true expert opinion is many times our very best assessment of informal education when we cannot hit it with methodologically precise instrumentation. So that, if there is a program that Nielsen's tells us is reaching X million kids, and it is also a program that this community of experts say is a damn good thing to put out there, then I think that the combination of those two facts is a powerful combination. From those two things we can reach some conclusion about the value of the program, and we can have some satisfaction that we know about the educational value of the program without ever having to run through the actual measurement of the psychological effects of the program, trying to meet some rigorous criteria that we cannot meet. I have been struggling, trying to form that concept for several years in my mind, and it is coming out for the first time here. It is that the community of expertise in combination with the multiplier effect of how many kids you reach provides a kind of measure of the opportunity you are offering kids. The quality times the quantity really is not a bad measure of opportunity.... The extent to which kids are realizing that opportunity is not known because we are so limited by our expertise in measurement--we are in a Stone Age in measurement. But we're not Stone Age in expertise.

Mark St. John: I find it interesting that those who have spent the most time actually trying to do assessment in this informal domain are the most humble--the least sure about what to do. Whereas it is the newcomers who are so confident of their methods.... I wonder if we aren't saying that experts are sometimes a very good surrogate for empirical measures...that they can bring to bear simultaneously a set of implicit, maybe even unconscious criteria, to their judgments...maybe they are the best instruments we have. There are similar other situations, like in hiring someone for a job, where we rely on interviews, subjective impressions, over more empirical measures because in our heart of hearts we don't trust them...the measures we have are so one-dimensional.

NEED FOR A MOSAIC OF STUDIES

Keith Mielke: You have to remember there are weaknesses, large weaknesses in all approaches. So that is why I like the concept of a mosaic--that's my term, a mosaic of studies--where the strengths of one design compensate for the weaknesses of another and, collectively, we have an array of studies that interlock. Within that array, a true experiment can be a very powerful addition.

I like the concept of an interlocking mosaic design. Any time we lay all our eggs in one basket we're in trouble. We know as the most fundamental of mass media effects propositions that people act in a social context, and so these things like Roger was talking about, the uses and gratifications literature, these are fundamental to informal learning. It's inextricably in that cauldron. I know that certain effects are not going to be visible at all in some of these designs that are akin to taking a snapshot from a satellite that's a million miles out in space. You can get a big picture but you miss an awful lot. I am reminded of the research we did on the follow-up of "Square One TV" where we had hours of interviews trying to track cognition by cognition on whether the problem was understood in a problem-solving segment, whether the mathematical principle in the problem-solving study was understood, and whether that could be extended to a different setting. That was a laborious data collection kind of thing and could not be touched with a 10-foot pole by other methodologies that would be longer term and broader in scope. So I think that there is a trade-off between sensitivity and the broadness of our view. Hence, again, we need the mosaic to compensate for that.

HOW TO ASSESS THE VALUE OF A CULTURAL OPPORTUNITY

Michael Templeton: There's always the tendency toward excessively high ego in these things.... You articulate goals and then make the mistake of thinking that you're in charge of the universe. In the informal domain the audience is in charge of the universe. The kids are choosing what to do. Given that, any success in this game is profound because you are actually competing with a hundred thousand alternatives. The funny thing is that you can hear somebody say, "We only reached 10 million people with that program," in a disappointed tone of voice because 40 million people watched some nitwit on TV. But 10 million people is a lot of people. I believe that we've always got to have a double value system: one value system for recognizing the worth of direct exposure and direct educational impact, and another value system that recognizes the inherent worth of the opportunity a resource represents--the opportunity that individuals have to make choices to participate. There's always a societal value in having created the opportunity by opening the door of the library, by opening the door to the museum, by having a TV show on 5 days a week in 90% of the markets in the country. That, by itself, is success. We jump too quickly over the assessment of success. The assessment of opportunity and choice is not quantified. From a value standpoint, we leap over that too quickly to get to the numerology that says, "If you got 10 million viewers, how come you didn't get 14?"

Mark St. John: The fact that a city has an opera is seen as important. It often doesn't matter how good it is--but the fact that there is an opera or symphony is seen as an important thing...there is value simply in its existence.... And note that we don't see evaluators worrying about how much people are learning at each opera.... Science museums are not judged by the same criteria. We tend to have much harsher, more formal educational criteria for these cultural resources.

Keith Mielke: Along those lines, and I cannot completely conceptualize this, I think informal learning frequently is not the sole cause of something. If we were thinking in analysis of variance terms, I would say it is not the primary variable responsible for the main effect. In the informal domain, what we do is that we put out resources that will interact with five other things in that person's life and then perhaps the combination has a long-term effect. Do you get credit for contributing to this interaction, even if you are not the main effect? Is there, in a strict accounting or experimental sense, an explicit or implicit requirement that you have to do the whole job yourself before you get any credit at all? I think that's an error, I think we ought to be quite explicitly joyful when we can do something that interacts with five or six other things and that in a multitude of small unknown, and perhaps unknowable, ways contributes to all of the diverse goals we have been talking about.

Mark St. John: Sometimes people will argue that informal learning leads to formal learning, and sometimes it is argued the other way around.... I see it more interactive than that--like a system of two coupled pendulums...or even more than two pendulums. The point is, you can put energy into the system through either pendulum...it will spread throughout the system...so it doesn't matter whether one gets excited about science in school or out of school...both reinforce each other... and both lead to this sense of being acculturated in science.

Michael Templeton. A question you could then ask is, "Is there a singular main outcome of informal science learning? Can you find a set of outcomes for which informal science experiences, if not a necessary condition, are at least a highly correlated requirement?" It's one thing to say that there are many subordinate outcomes that occur as a consequence of informal learning. And there is a fair amount of rhetoric about the things informal learning environments do that other settings can't or don't do very well. But I'm not aware of any data or any analysis that comes at all close to the assertion that there is a main outcome, a main effect for informal learning. It may be that informal science experiences serve as an important adjunct activity, or it may be that we do not know how to find the main effect. In other words, are we looking for neutrinos, or for phlogiston?

THE NEED NOW IS FOR A VARIETY OF ASSESSMENT EXPERIMENTS

Roger Miles: If you look at the reasons people go to the mass media, the sorts of satisfactions or gratifications they seek, the answers fall into four large categories. There is learning and knowledge. There's personal identity--to get a better understanding of yourself, to reinforce your own view of yourself. Then there is integration and social interaction--to be with people, to give you something to talk about with people, to maintain contact with family and friends. And lastly, there's leisure--simply to kill time pleasurably, take your mind off your worries, etc. Surely people use museums for all of those reasons and they're not mutually exclusive. One has to take into account the complete picture. When we are thinking about the role of a museum or science center in informal science education, we ought to think about museums as providing a sort of framework to which people come and can engage in various social functions and which also provide an opportunity for learning to take place. If you're going to do any evaluation, surely the most profitable way is to look at the resources and setting provided and think about the potential, the opportunity--is it there? Because to try and measure individual learning outcomes is actually difficult, and the approach is not particularly applicable.

Bret Waller: Is it true that you can throw out the last three functions--that all of the nonlearning outcomes are wasted time and money so far as NSF is concerned? Is it only the learning part that counts and we ought to discourage all those other things because they're just taking up time and space as well? But I think we need to deal with the whole package and look at it as a whole--we need to recognize they are inextricably connected--that we need to recognize that what we learn has a whole lot to do with your self-image. These are intimately connected. If you take that perspective, then in thinking about assessment you have to be prepared to deal a lot more with the whole phenomenon of museums, than if you just focus so single-mindedly on the learning outcomes.

APPLIED RESEARCH

Michael Templeton: My sense is that most of the issues we've talked about lie in the domain of applied research. A much smaller fraction of assessment issues lie in the area where evaluation measures come readily to hand. It seems clear that even in areas where the intent is pragmatic assessment, there is a research level required to design the instrument, to design the means to answer the question. This is a field that does not have standardized measures or instruments. More substantially, most of the things that have been talked about call for applied research, where significant thinking is necessary before you begin. I say applied research because NSF itself has mostly supported pure research in teaching and learning. Applied education research is more difficult to get support for. This discussion reinforces the need to solve that problem. One of the characteristics of research in informal education is that most research methodologies that generate samples and averages are of limited utility for an environment characterized by diversity, heterogeneous audiences, subaudiences, and subcomponents. There has to be a diversity of means, measures, and examples or most of the simple information we could get is lost.

There is also a need for some overarching view of what the nature of the public understanding of science is: a need for an overarching view of what the body of informal science education is. The field itself has a very primitive notion of its own philosophy. We need more philosophical discourse and a good deal more theoretical thinking. There needs to be more strategy from a societal standpoint about where informal science education is positioned. At the level of how informal learning functions, a much wider variety of paradigms is needed. One of the things I've been struck with in this dialogue is how inarticulate we all are, because we don't have very good visual, or verbal, or other models to describe the transactions and interactions of informal learning. We are imagery poor; we are paradigm poor.

Finally, there need to be more meta-analyses--attempts to synthesize the best of the work that has been done, combined with available quantitative, robust information to draw more general conclusions. If you will, a strategic assessment based on a selective look at prior research and prior data. This adds up to a substantial recipe for better understanding. I'm not sure it satisfies an internal NSF interest in the machinery for a report card on how well its programs are doing. But it does rather well in suggesting some directions and some boundaries for the domain of informal science education.

SRI International
333 Ravenswood Avenue
Menlo Park, California 94025-3493
(415) 326-6200
TWX: 910-373-2046
Telex: 334 486